JOURNAL OF APPLIED PHYSICS

JOURNAL OF APPLIED PHYSICS

Vol. 16, No. 11 NOVEMBER, 1945

In This Issue

		Page
Proposed "Atomic Energy Act of 1945"		645
Gas Flow in the Mass Spectrometer	Richard E. Honig	646

Contributed Original Research

Use of Direct Current in Induction Electron Accelerators	W. F. Westendorp	657
Crucible Failure in the Induction Melting Process	Andrew Gemant and Joseph Sticher	661
Double-Solenoid Coil System for Magnetic Measurements	E. W. Frowe and C. J. Aronson	667
Class of Electron Lenses which Satisfy Newton's Image Relation	R. G. E. Hutter	670
Rigorous Treatment of the Electrostatic Immersion Lens Whose Axial Potential Distribution is Given by: $\phi(z)=\phi_0 e^{K \arctan z}$	R. G. E. Hutter	678
Methods of Specifying the Properties of Viscoelastic Materials	Turner Alfrey and Paul Doty	700
Double-Yoke Plate Permeameter	William F. Brown, Jr. and Sidney M. Rubens	713
Electron Diffraction Camera for the Study of High Temperature Surface Reactions	Earl A. Gulbransen	718
Virtual Masses of Rectangular Plates and Parallelepipeds in Water	Yee-Tak Yu	724
Electron Microscopical Replica Techniques for the Study or Organic Surfaces	R. Bowling Barnes, Charles J. Burton, and Robert G. Scott	730
Stability of the Low Pressure Mercury Arc as a Function of Area at the Cathode Pool	Carroll W. Lufcy and Paul L. Copeland	740

In Every Issue

Here and There	655
New Booklets	656
Index to Advertisers	xxviii

BOARD OF EDITORS

Elmer Hutchisson, *Editor*

Associate Editors

R. B. Barnes
Otto Beeck
W. F. Busse
S. H. Caldwell
Clark Goodman
Selig Hecht
W. J. Lyons
Morris Muskat
Thomas H. Osgood
Paul E. Sabine
Frederick Seitz
J. R. Townsend

Publication of the **American Institute of Physics**

ADMINISTRATIVE STAFF

Henry A. Barton, *Director*
Margaret S. Griffin, *Publications Manager*
Theodore Vorburger, *Business Manager*
John T. Tate, *Adviser on Publications*

GOVERNING BOARD

Paul E. Klopsteg, *Chairman*
Wallace Waterfall, *Secretary*
George B. Pegram, *Treasurer*
T. D. Cope
R. B. Dow
L. A. DuBridge
Floyd A. Firestone
Harvey Fletcher
R. C. Gibbs
K. S. Gibson
Hermann Mark
A. G. Worthing
John T. Tate
L. W. Taylor
N. W. Taylor
Wallace Waterfall

The Journal of Applied Physics, published monthly at Prince and Lemon Sts., Lancaster, Pa., is devoted to physics in its role as the science basic to other natural sciences and to the arts and industries. Previous to 1937 this publication was known as Physics. The Journal of Applied Physics publishes editorials, news of physicists, and reviews, as well as technical papers of applied physics.

Manuscripts should be submitted to Elmer Hutchisson, Editor, Case School of Applied Science, Cleveland 6, Ohio.

Subscriptions and orders for back numbers should be addressed to Prince and Lemon Streets, Lancaster, Pa. or to the American Institute of Physics, 57 East 55 Street, New York 22, N. Y.

Subscription Price

	U. S. and Canada	Elsewhere
To members of scientific societies	$5.00	$5.70
To all others	7.00	7.70

Back Number Prices

Complete set: Vol. 1, July, 1931–Vol. 15, 1944—$94.50
Yearly back number rate: $7.70
Single copies: $0.70 each.

Changes of address and proofs should be addressed to the Publications Manager.

Advertising rates supplied on request. Orders, advertising copy, and cuts should be sent to the American Institute of Physics.

Entered as second class matter January 22, 1937, at the Post Office at Lancaster, Pennsylvania, under the Act of March 3, 1879
Accepted for mailing at the special rate of postage provided for in the Act of February 28, 1925, authorized May 2, 1932.

Copyright 1945, by the American Institute of Physics.

Journal
of
Applied Physics

Volume 16, Number 11 November, 1945

The Proposed "Atomic Energy Act of 1945"

"*PLENARY supervision and control, so far as the jurisdiction of the United States extends, over all sources of atomic energy and over all matters connected with research on the transmutation of atomic species, the production of nuclear fission, and the release of atomic energy.*"

That, in short, is the scope of the May-Johnson Bill introduced into both Houses of Congress on October 3 (S 1463 and HR 4280). Under this Bill the President would appoint a nine-man Atomic Energy Commission which would set up its own Administrator, Deputy Administrator, and necessary staff and working facilities. The powers granted the Commission and the Administrator are totally unprecedented in a democratic form of society during peacetime. Under it there would be transferred into the custody and control of the Commission all stocks of thorium and uranium and their ores, all plants, facilities, equipment, and materials for the refining, production, or utilization of thorium, uranium, or elements higher than uranium in the periodical table; and all processes, technical information, contracts, agreements, leases, patents, patent applications, and discoveries relating to thorium or uranium production or utilization. Then, at times set by the Commission or the Administrator, all persons having any right, title, interest, or claim in such property or processes or any ores from which such materials may be produced must declare these in the form and to the extent prescribed by the Commission or the Administrator.

Under this Bill the Administrator "is authorized to conduct research and experimentation in the field of nuclear fission, the transmutation of atomic species, and closely related phenomena, and to proceed with the development of any and all processes or methods for the release of atomic energy, and for the exploitation and use thereof for military, industrial, scientific, or medical purposes: *provided, however*, that it shall be the policy of the Commission and of the Administrator, in accord with the objectives of this Act, to utilize, encourage, and aid colleges, universities, scientific laboratories, hospitals, and other governmental, nonprofit, or private institutions equipped and staffed to conduct research and experimentation in this field." Further, however, the Bill provides that "it shall be unlawful for any person to conduct research or experimentation involving the release of atomic energy in amounts deemed and to be prescribed by the Commission as constituting a national hazard or being of military or industrial value, without the consent of the Administrator and upon such conditions as the Administrator, with the approval of the Commission, may prescribe." To soften this provision there is added, "In administering this provision, the Commission and the Administrator shall interfere to the least possible extent with small-scale experimentation in research laboratories of nonprofit institutions."

Under "Security Regulations" the Bill states, "The Commission is authorized and directed to establish, and to provide for the administration of, security regulations governing the collection, classification, dissemination, publication, transmission, handling, and communication" of information and material in this field. The penalties for security violations are extraordinary, amounting to as much as a fine of $300,000 as well as imprisonment for 30 years.

It is inconceivable that a bill which would throttle scientific research in the manner here proposed would receive serious consideration by the Congress. As Dr. K. T. Compton has said, supposing that when the automotive engine was in its infancy, similar secrecy restrictions had been imposed because of its possible military applications; in such a case, one of our greatest assets—our overwhelming superiority in mechanical equipment—would have been completely lacking in the present war. It behooves every scientist to read this Bill carefully and to use his influence in keeping open all channels that lead to progress in science.—E. H. (Nov. 1, 1945)

Gas Flow in the Mass Spectrometer

RICHARD E. HONIG
Physics Department, Massachusetts Institute of Technology, Cambridge, Massachusetts
(Received May 21, 1945)

If a mass spectrometer is to be used successfully for gas analyses, the gas flow through the system must meet a number of requirements. The three modes of gas flow—molecular, intermediate, and viscous—are reviewed, and their working ranges are discussed in terms of the pressure of the system. The upper limit for molecular flow, through capillaries or through small openings, is found to lie at a pressure where the mean free path of the molecules is between 15 and 40 times the diameter of such a gas leak. Expressions are developed to show that, for molecular flow, the ion beam intensity is independent of molecular weight but is a linear function of the sample pressure, and that the principle of linear superposition applies to mixture peaks. Since the composition of a gas mixture in a reservoir is found to change with time (a numerical example is worked out), the necessity for analyses of short duration becomes evident. The practical aspects of designing the gas system for molecular flow and for efficient ion production are discussed in detail. The various flow requirements are best met by a gas leak consisting of a small hole in a diaphragm, or several holes in parallel. A relatively high gas pressure within the electron beam boundaries and resulting large ion currents are produced by introducing the gas as a molecular beam into a gas-tight ionizing region.

INTRODUCTION

WHILE the design, construction, and operation of mass spectrometers have received considerable attention,[1] studies of the gas flow in these instruments have not been published. At first, the spectrometer was used mainly to measure isotopic abundance ratios[2] and to study the products of ionization and dissociation by electron impact.[3] Recently its use has been extended to gas analyses,[3,4] and it is in this application that a knowledge of gas flow is of paramount importance. An instrument of the Nier type[5] used for gas analyses is shown in Fig. 1.

To perform satisfactory gas analyses with the mass spectrometer, the following primary requirements should be fulfilled. (1) The ion beam intensity should be directly proportional to the pressure of the gas in the sample, but should be independent of its molecular weight. (2) In a gas mixture, the presence of one component should not affect the peak intensities due to the other components, i.e., all mixture peaks are formed by the linear superposition of individual intensities.

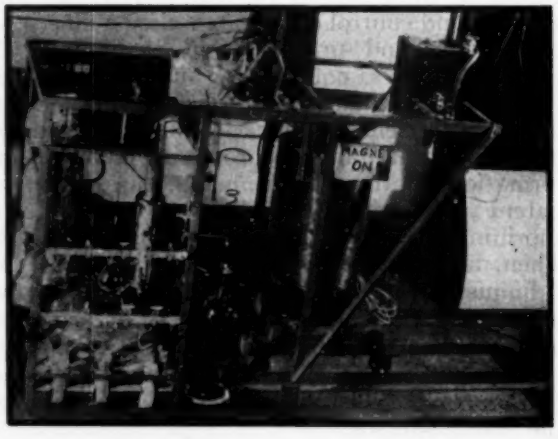

FIG. 1. Photograph of M.I.T. spectrometer* used for gas analyses on American Petroleum Institute Project 43c ("Studies of the Effect of Radioactivity on the Transformation of Marine Organic Materials into Petroleum Hydrocarbons").

[1] For a review article, including developments until 1942, see: E. B. Jordan and L. B. Young, "A Short History of Isotopes and the Measurement of Their Abundances," J. App. Phys. 13, 526–538 (1942).

[2] See for example: A. O. Nier, "A Mass-Spectrographic Study of the Isotopes of Hg, Xe, Kr, Be, I, As, and Cs," Phys. Rev. 52, 933–937 (1938).
A. O. Nier, "The Isotopic Constitution of Strontium, Barium, Bismuth, Thallium, and Mercury," Phys. Rev. 54, 275–278 (1938).
Review Article: D. Rittenberg, "Some Applications of Mass Spectrometric Analysis to Chemistry," J. App. Phys. 13, 561–569 (1942).

[3] J. A. Hipple, "Gas Analysis with the Mass Spectrometer," J. App. Phys. 13, 551–560 (1942).

[4] H. Washburn, D. D. Taylor, E. E. Hoskins, and R. V. Langmuir, "Application of Mass Spectrometry to Analysis of Hydrocarbon Gas Mixtures," Transactions of the California Natural Gas Association, Fall Meeting (1940).
Herbert Hoover, Jr. and Harold Washburn, "Analysis of Hydrocarbon Gas Mixtures by Mass Spectrometry," California Oil World (Nov., 1941).
H. W. Washburn, H. F. Wiley, and S. M. Rock, "The Mass Spectrometer as an Analytical Tool," Ind. Eng. Chem. Anal. Ed. 15, 541–547 (1943).
H. W. Washburn, H. F. Wiley, S. M. Rock, and C. E. Berry, "Mass Spectrometry," Ind. Eng. Chem. Anal. Ed. 17, 74–81 (1945).

[5] A. O. Nier, "A Mass Spectrometer for Routine Isotope Abundance Measurements," Rev. Sci. Inst. 11, 212–216 (1940).

* Constructed by author in partial fulfillment of the requirements for the degree of Doctor of Philosophy from the Massachusetts Institute of Technology.

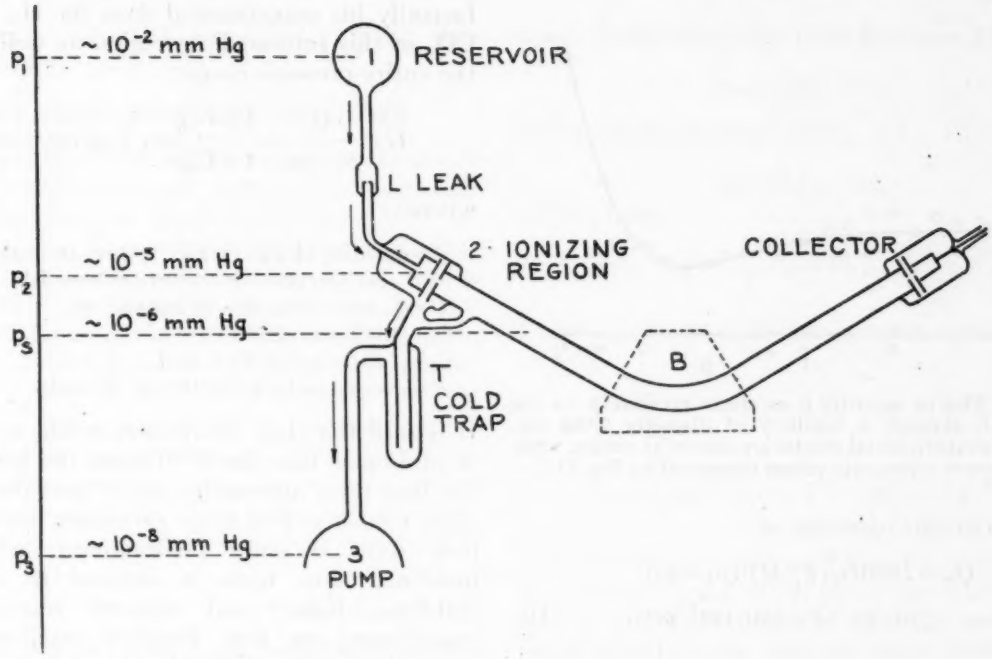

FIG. 2. Essential components of a spectrometer gas system. Approximate pressure levels of various parts are indicated by dotted lines.

(3) The gas flow, and as a result the ion beam intensity, should remain essentially constant during an analysis.

All of these requirements can essentially be fulfilled if *molecular* flow through a properly designed gas system is established by the correct choice of pressures. Figure 2 is a schematic diagram of the essential components of a spectrometer gas system. Gas contained in reservoir 1 at pressure p_1 flows through a suitable leak L into the ionizing region 2 of the spectrometer proper. In this ionizing region, which is maintained at pressure p_2, excitation, ionization, and dissociation of the gas molecules are produced by electron impact. A certain fraction of the ionized molecules and fragments, after being accelerated by adjustable electric fields, is collimated by slits to form an ion beam and emerges into the tube proper. The ion beam is analyzed by magnetic field B, and the components of the resulting mass spectrum are measured electrically at the collector end. The neutral molecules and fragments are removed, through cold trap T, by diffusion pump 3 which is at pressure p_3. If p_1 be chosen so low that the corresponding mean free path of the molecules is much larger than the effective diameter of the leak opening, then the probability of intermolecular collisions within the leak is negligibly small compared to the probability of molecules colliding with the walls. The condition that each molecule proceeds through the leak individually is known as *molecular flow*. In Fig. 2 the orders of magnitude shown for pressures p_1, p_2, p_s, and p_3 represent average values which will permit molecular flow as well as proper operating conditions for the spectrometer tube.

LAWS OF GAS FLOW

The laws of molecular flow were first derived by Knudsen in his classical work on the flow of rarefied gases through tubes and small openings.[6,7] Defining Q_m as the time rate of change of the pressure-volume product which, at unit pressure, becomes equal to the volume V of gas discharged per second by a capillary tube or by a small opening, the Knudsen equation may be written for cylindrical tubes as:

$$Q_m = d(\bar{p}V)/dt = 3800(D_T^3/L)(T/M)^{\frac{1}{2}}(p_1 - p_2)$$
$$= m_T(p_1 - p_2) \text{ bar-cm}^3/\text{sec.}, \quad \text{(1a)}$$

[6] M. Knudsen, "Molekularströmung und innere Reibungsströmung," Ann. d. Physik [IV] **28**, 75–131 (1909).
[7] M. Knudsen, "Die Molekularströmung der Gase durch Öffnungen und die Effusion," Ann. d. Physik [IV] **28**, 999–1016 (1909).

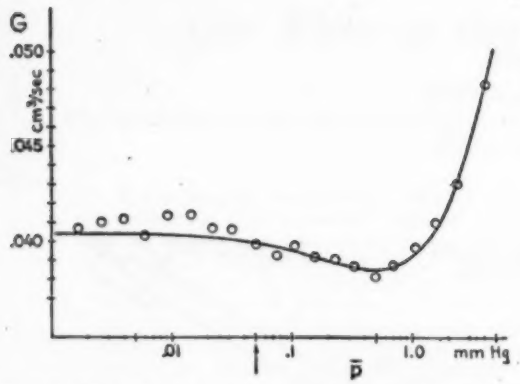

FIG. 3. Plot of quantity G *vs.* mean pressure $\bar{p}$ for the flow of O_2 through a capillary of diameter 0.066 cm. Knudsen's experimental results are shown as circles, while the solid curve represents values computed by Eq. (3).

and for circular openings as:

$$Q_m = 2860 D_0^2 (T/M)^{\frac{1}{2}}(p_1 - p_2) = m_0(p_1 - p_2) \text{ bar-cm}^3/\text{sec.}, \quad (1b)$$

where

D_T = diameter of tube in cm,
D_0 = diameter of opening in cm,
L = length of tube in cm,
T = absolute temperature in °K,
M = molecular weight (scale: $O^{16} = 16$),
p_1 = pressure at high side of tube or opening in bars,
p_2 = pressure at low side of tube or opening in bars,
$\bar{p} \equiv (p_1 + p_2)/2$ = mean pressure in bars,
m_T = coefficient of molecular flow through tubes, defined by Eq. (1a), and
m_0 = coefficient of molecular flow through openings, defined by Eq. (1b).

In deriving Eqs. (1a) and (1b) above, it is necessary to assume that the mean free path L_M of the molecules is large compared to the tube or opening diameter. Knudsen found experimentally that Eq. (1a) is applicable in the low pressure range provided the mean free path, corresponding to the mean pressure of the system, is greater than about $30\,D_T$. At high pressures ($L_M \ll D_T$) the gas flow Q_v is viscous and follows Poiseuille's law:

$$Q_v = (\pi D_T^4/128\eta L)\bar{p}(p_1 - p_2) = v\bar{p}(p_1 - p_2) \text{ bar-cm}^3/\text{sec.}, \quad (2)$$

where η = coefficient of viscosity in poises, and v = coefficient of viscous flow, defined by Eq. (2).

No simple relation exists for the intermediate pressure range ($L_M \cong D_T$). However, Knudsen was able to establish the following semi-empirical relation, which, by the proper choice of the two constants C_1 and C_2, could be made to fit satisfactorily his experimental data for H_2, O_2, and CO_2 in this intermediate region as well as over the entire pressure range:

$$G \equiv \frac{Q}{p_1 - p_2} = \frac{1 + C_1\bar{p}}{1 + C_2\bar{p}} m_T + v\bar{p} \text{ cm}^3/\text{sec.}, \quad (3)$$

where

G = quantity of gas ($= \bar{p}V$) flowing through the tube per sec. per unit pressure differential,
$Q = Q_m + Q_v$ = total flow in bar-cm³/sec.,
$C_1 = (1.00 D_T/\eta)(M/RT)^{\frac{1}{2}}$,
$C_2 = (1.23 D_T/\eta)(M/RT)^{\frac{1}{2}}$, and
R = gas constant $= 8.31 \times 10^7$ erg/°K mole.

It is seen that Eq. (3) reduces to the coefficient of molecular flow for $\bar{p} \to 0$ since the fraction in the first term approaches unity and the second term vanishes. For large pressures, the viscous flow term becomes predominant, while the molecular flow term is reduced to $m_T C_1/C_2 \cong 0.8\, m_T$. Klose[8] and Adzumi[9] repeated the experiments on flow through capillaries and generally confirmed Knudsen's results, although Adzumi's value of the molecular flow term is closer to $0.9\, m_T$ for pure gases and is as low as $0.6\, m_T$ for certain gas mixtures.

An interesting graph is obtained by plotting the quantity G as a function of the mean pressure $\bar{p}$. Figure 3 shows Knudsen's results for O_2 flowing through a capillary. The region of molecular flow is represented by the horizontal line, intermediate flow by the minimum region and viscous flow by the steep slope on the right. If it be arbitrarily assumed that the region of molecular flow extends to the point where the quantity G has dropped 1 percent below its constant value ($= m_T$) (see arrow in Fig. 3), this upper limit of pressure $\bar{p}_{max}$ is such that the minimum mean free path $(L_M)_{min} = 30\, D_T$. This value of L_M lies between $15\, D_T$ and $40\, D_T$ when Knudsen's results for H_2 and CO_2 and Azdumi's for H_2, C_2H_2, C_3H_6, and their mixtures are plotted. Thus the inequality $L_M > D_T$, yielding molecular flow as mentioned above, is numerically quite well defined for pure gases as well as gas mixtures. Knudsen's results[7] for gas flow through openings do not give rise to a minimum region of G for intermediate values of $\bar{p}$, but here again the region of molecular flow is clearly

[8] W. Klose, "Über die Strömung verdünnter Gase durch Kapillaren," Ann. d. Physik [V] 11, 73–93 (1931).
[9] H. Adzumi, "Studies on the Flow of Gaseous Mixtures through Capillaries. III. The Flow of Gaseous Mixtures at Medium Pressures." Bull. Chem. Soc. Japan 12, 292–303 (1937).

defined and extends to a $\bar{p}_{max}$ such that $(L_M)_{min} \geqq 20_D$.

GAS FLOW IN THE SPECTROMETER

The assertion that all the requirements stated in the introduction are satisfied by molecular flow must now be proved. To show that for molecular flow the ion beam intensity is directly proportional to the reservoir pressure p_1 and is independent of molecular weight (requirement (1)), it is first assumed that the intensity depends linearly on the number of ions formed per second. This number, in turn, is proportional to the density of gas molecules in the ionizing region, and therefore to pressure p_2. Since, for equilibrium, the gas flow into the ionizing region (Q_{12}) must equal the flow out (Q_{23}), one may write, neglecting the presence of the cold trap T:

$$Q_{12}=(T/M)^{\frac{1}{2}}(p_1-p_2)/r_{12}=Q_{23}$$
$$=(T/M)^{\frac{1}{2}}(p_2-p_3)/r_{23},$$

where r_{12}, r_{23} = coefficients including the line resistances between 1 and 2, and 2 and 3. But $p_1(\geqq 10^{-2}\text{ mm}) \gg p_2(\geqq 10^{-5}\text{ mm}) \gg p_3(\geqq 10^{-8}\text{ mm})$. Therefore, to a good approximation:

$$p_2=p_1 r_{23}/r_{12}. \tag{4}$$

Equation (4) indicates that the gas pressure in the ionizing region, and, therefore, the ion beam intensity, is directly proportional to the reservoir pressure and depends only on the geometry of the lines, but is independent of molecular weight.

By generalizing the basic condition for molecular flow stated above, it can be shown that in a gas mixture the different components do not influence each other's flow (requirement (2)). If it be true that, within the leak boundaries, the progress of any one molecule is not affected by other molecules present in the reservoir, it follows that the nature of these other molecules is immaterial. Therefore the molecules of one species present in a gas mixture will proceed through a leak in a manner independent of any other components, and their discharge rate is a function only of their own partial pressure differential and molecular weight, and of the leak geometry—provided, of course, that the sum of all partial mean pressures lies below the upper limit for molecular flow. Consider a gas mixture consisting of components $a, \cdots i, \cdots n$ with partial reservoir pressures $p_{1a}, \cdots p_{1i}, \cdots p_{1n}$, whose total amount in the reservoir

$$p_1 V_1 = V_1 \sum_a^n p_{1i} = p_1 V_1 \sum_a^n f_i,$$

where f_i represents the mole fraction. Then

$$p_{1a}=f_a p_1, \text{ etc.},$$

and similarly, because of Eq. (4)

$$p_{2a}=f_a p_2, \text{ etc.},$$

which yields:

$$p_2=\sum_a^n p_{2i}=p_2 \sum_a^n f_i. \tag{5}$$

Therefore the total discharge rate may be written as

$$Q=\sum_a^n Q_i=(T^{\frac{1}{2}}/r_{12})\sum_a^n (p_1-p_2)_i/M_i^{\frac{1}{2}},$$

or

$$G=Q/(p_1-p_2)=(T^{\frac{1}{2}}/r_{12})\sum_a^n f_i/M_i^{\frac{1}{2}}. \tag{6}$$

Experiments have shown[10] that for a fixed electron energy and a given gas a the intensity I_a^+ of any one of the peaks in its mass spectrum is proportional to the electron beam intensity I^- and the concentration of molecules in the ionizing region, and depends on the kind of gas:

$$I_a^+=kI^- p_{2a} S_a, \tag{7}$$

where k = a proportionality factor and S_a = specific ionization, characteristic of gas a. Since in a gas mixture the components contribute linearly to the pressure in the ionizing region (Eq. (5)), it appears that the total intensity of any mixed peak

$$I^+=kI^-\sum_a^n p_{2i}S_i=kI^- p_1(r_{23}/r_{12})\sum_a^n f_i S_i. \tag{8}$$

If the specific ionization numbers $S_a, \cdots S_i, \cdots S_n$ are determined experimentally from Eq. (7) by running each component gas separately, a mixture of n components can be completely analyzed by applying Eq. (8) n times. Analysis by this procedure has been described in detail by Washburn and his collaborators.[4]

Constancy of gas flow is the third main requirement for performing proper gas analyses. By making the reservoir volume V_1 sufficiently large, the change in reservoir pressure Δp_1 occurring during the course of an analysis may readily be kept down to the order of 1 percent or less. Thus the changes in ion beam intensity due to changes in gas pressure will be of the same

[10] See for example: D. P. Stevenson and J. A. Hipple, Jr., "Ionization and Dissociation by Electron Impact: Normal Butane, Isobutane, and Ethane," J. Am. Chem. Soc. 64, 1588–1594 (1942).

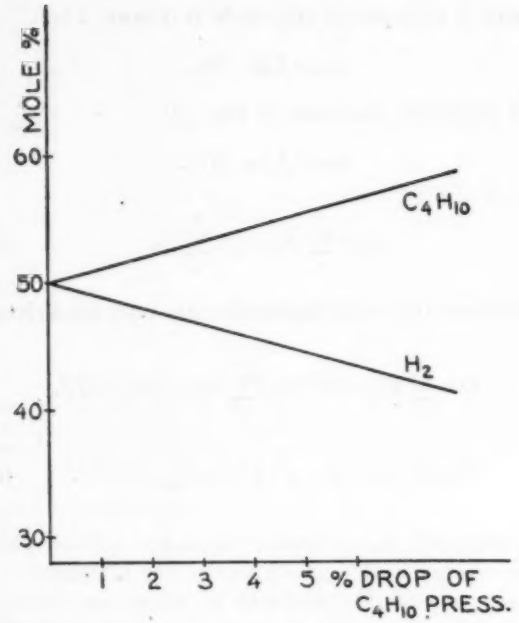

FIG. 4. Plot showing the mole percent in a butane-hydrogen mixture (initially 1:1) *vs.* the percent drop in butane pressure.

order as the accuracy with which the peaks are measured. However, a far more serious consideration arises from the fact that the rate of molecular flow of a gas depends inversely on the square root of its molecular weight (Eq. (1)). While it is true that the ion beam intensity is independent of molecular weight (see Eqs. (4) and (7)), the fact remains that the composition of a gas mixture in the reservoir will change slowly due to the different flow rates of its components. To express this change mathematically, consider a reservoir of volume V_1, containing initially an amount of gas p_0V_1 which after a time is reduced to p_1V_1. Since the rate of gas lost from the reservoir equals the rate of flow through the system, one may write at time t:

$$-d(p_1V_1)/dt=-V_1dp_1/dt=Q_t$$
$$=(T/M)^{\frac{1}{2}}(p_1-p_2)/r_{12}.$$

But since $p_2=p_1r_{23}/r_{12}$,

$$-V_1dp_1/dt=(T/M)^{\frac{1}{2}}(r_{12}-r_{23})p_1/r_{12}^2=C'p_1/M^{\frac{1}{2}}$$

and

$$dp_1/p_1=-C'dt/V_1M^{\frac{1}{2}}=-Cdt/M^{\frac{1}{2}},$$

which yields on integration

$$\ln(p_1/p_0)=-Ct/M^{\frac{1}{2}}$$

or

$$p_1=p_0\exp(-Ct/M^{\frac{1}{2}}). \quad (9)$$

Applying Eq. (9) to a binary mixture of components a and b, one obtains:

$$\frac{\ln(p_1/p_0)_a}{\ln(p_1/p_0)_b}=\left(\frac{M_b}{M_a}\right)^{\frac{1}{2}}$$

or

$$(p_1/p_0)_a=(p_1/p_0)_b^{(M_b/M_a)^{\frac{1}{2}}}. \quad (10)$$

To take a rather extreme case, suppose the mixture consists initially of equal parts of butane ($M_B=58$) and hydrogen ($M_H=2$), and it is desired to determine its composition after a certain interval. The reservoir pressure of hydrogen corresponding to, say, a 1 percent decrease in butane pressure is computed from Eq. (10) as

$$(p_1/p_0)_H=0.990^{(58/2)^{\frac{1}{2}}}=0.947,$$

from which the new composition is obtained as

$$\frac{0.990}{0.990+0.947}100=51.1 \text{ percent butane}$$

and

$$\frac{0.947}{0.990+0.947}100=48.9 \text{ percent hydrogen.}$$

These changes show up very plainly in Fig. 4 where, over the region considered, the mole percent of the components changes linearly with the butane pressure drop. To perform an analysis accurate to ± 0.5 percent, it is therefore necessary to limit $(\Delta p_1/p_0)_B$ in this case to about 0.5 percent. While this example represents one of the most unfavorable cases encountered, it points out clearly the necessity of large reservoirs and analyses of short duration.

It is interesting to note here that the enrichment in heavier components occurring in the reservoir for molecular flow is not encountered with pure viscous flow. Equation (2) shows that such flow is independent of molecular weight, but is inversely proportional to the viscosity coefficient η, a quantity which can only be determined by experiment. It turns out that in the case of a gas mixture the components flow out together at a rate determined by a new coefficient. Not only is this coefficient characteristic of the mixture, but its value depends in an entirely unpredictable manner on the relative abundances of the components.[9,11] Since the

[11] W. Schudel, "Kinetische Gastheorie und Viskosität binärer Gasmischungen," Monatsbulletin Schweiz. Gas- und Wasserfach **22**, No. 2, 21–31 (1942).

principle of linear superposition does not apply in this case, it is not possible to perform analyses in the fashion outlined above. Furthermore, it must be realized that even though pure viscous flow could easily be maintained between reservoir and ionizing region, the flow between ionizing region and pump would of necessity be molecular. Therefore, enrichment in heavier components would still occur in the spectrometer.

DESIGN OF GAS SYSTEM

Besides the major requirements for proper gas flow discussed in detail above, there are a number of practical considerations that are encountered when the actual design of a gas system is undertaken. For a complete gas analysis, provisions must be made to admit in turn to the reservoir the gas mixture and all the necessary calibrating gases at a pressure that is known and can be readily controlled. Figure 5 shows in diagram form a satisfactory gas-handling system, adapted from a design published by P. T. Smith *et al.*[12] Gas from any one of the bottles is admitted into the manifold to the right of stopcock S_3 at a pressure measured on the manometer. The known amount of gas contained in the gas pipet between S_2 and S_3 is then expanded into reservoir 1. After the measurement has been completed, the reservoir and gas-handling system are evacuated through S_4.

The physical dimensions of the gas system are determined by the following considerations. The cross section of the manifold should be as small as possible to minimize dead space, yet not so small that appreciable composition changes in the unknown can occur as it flows from the sample bottle into the pipet. An interior diameter of 3 to 4 mm, corresponding to about 1 cm³ per 10 cm of line, appears satisfactory. Stopcock S_1 serves to cut off the manifold when unknown samples are run. A reservoir volume, V_1, of 10 liters is sufficiently large to keep the gas pressure drop at about 2 percent per hour when used with an appropriate leak. If the gas pipet volume V_s is made 4 cm³ and filled to a pressure of 10 cm Hg, measurable on the manometer to better than one part in 100, the initial reservoir pressure is given by Boyle's law as $p_0 = 4 \times 10^{-3}$ cm Hg or 40 microns. Then the mean pressure $\bar{p} = 20$ microns, and it follows that the corresponding mean free path L_M is about 0.2 cm for CO_2 and somewhat higher for most other gases. Thus, to insure molecular flow, the leak diameter should be smaller than about 0.01 cm. This limitation suggests the use of an opening, since capillaries are not usually available in such small sizes. With the upper limit fixed, the actual diameter is determined by the gas flow desired. This flow should be large enough to yield ample peak inten-

[12] P. T. Smith *et al.*, "A High Sensitivity Mass Spectrograph with Automatic Recorder," Rev. Sci. Inst. **8**, 51–55 (1937).

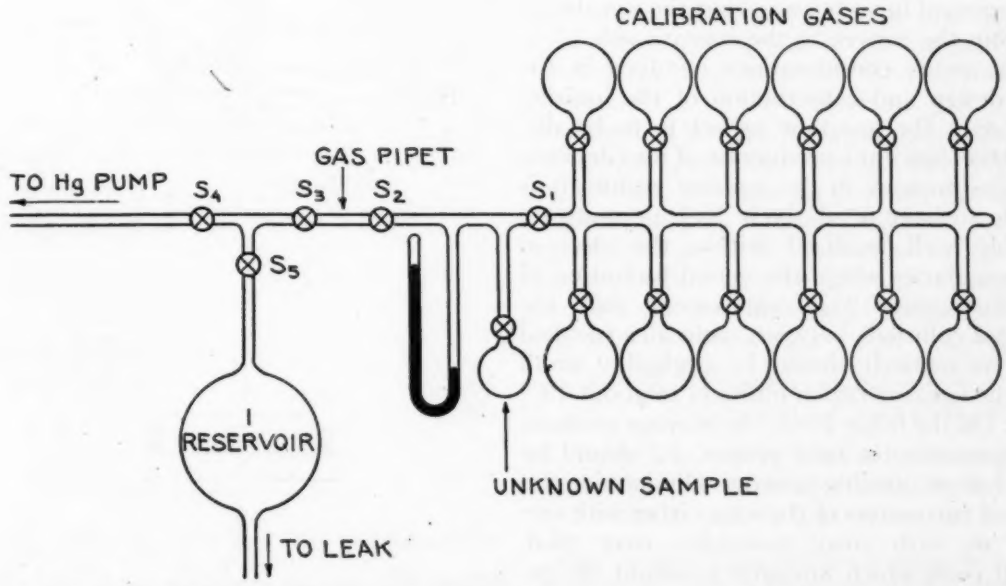

FIG. 5. Diagram of gas-handling system. Stopcock S_1 serves to minimize dead space when small amounts of unknown samples are run. Reservoir 1 is made up of two five-liter bottles.

FIG. 6. Enlarged cross-sectional view (not drawn to scale) of ordinary 7-mm Pyrex tube serving as diaphragm for opening of 1-mil diameter. Note how wall thickness is tapering off toward hole.

sities above background, yet not be too large in order to keep the reservoir pressure drop down to a minimum. It turns out that, together with the reservoir pressure and volume specified above, an opening about 1 mil (=0.0025 cm) in diameter satisfies all requirements. If in certain cases the diameter limited by molecular flow should not satisfy the discharge requirements, it is an easy matter to arrange a number of small-diameter leaks in parallel—a device first employed by Knudsen.[6]

Although it would be difficult to make an opening in an *ideal* diaphragm, whose wall thickness is negligible compared to the hole diameter, small openings can be obtained very easily in the following manner: ordinary thin-walled glass tubing is heated at one end in a flame and allowed to shrink while at the same time air is blown through. By judicious heating and blowing, openings of any desired diameter are obtained, of a shape as shown in Fig. 6. While the diaphragm is neither plane nor uniform in wall thickness, it serves very well in actual practice. In order to keep the hole from plugging, the leak is oriented in such a way that the gas always flows from the convex to the concave side.

Of the many considerations involved in the proper design and construction of the ionizing region, only the gas flow aspect is to be discussed here. Since the production of ions depends on the gas pressure in the ionizing region, it is desirable to have a relatively high pressure p_2, preferably well localized within the electron beam boundaries where the actual formation of ions takes place. The requirement that the number of collisions between molecules (neutral as well as ionized) should be negligibly small within the ionizing region limits p_2 to about 10^{-4} mm Hg. On the other hand, the average pressure in the spectrometer tube proper, p_s, should be kept as low as possible in order to minimize the number of encounters of the ions, either with one another or with stray molecules, over their extended path which amounts to about 60 cm in the case of the 60° Nier type instrument. In practice, p_s may be about 10^{-6} mm Hg. To obtain this pressure differential (p_2-p_s), two different methods are utilized. First, the average pressure in the ionizing region is raised by inserting a high flow resistance between the region and the rest of the tube. Secondly, the local pressure within the electron beam boundaries can be boosted further by introducing the gas as a well-defined intense molecular beam.

A relatively gas-tight ionizing region is obtained by the simple construction indicated in Fig. 7 (a), where the narrow gap between side walls and bottom is needed for electrical insulation. Gas-tightness can be improved by insulating spacers as shown in Fig. 7 (b), in which case the gas molecules can only escape through the electron and ion slits. However, such insulating spacers must be mounted so they cannot "see" the ion beam; otherwise the field distortion resulting from stray charges collecting on them may completely suppress the beam. If the spectrometer is to be used for hydrocarbon research, it is important that the effects of thermal cracking of gases at the hot filament should be minimized. To prevent the products of thermal cracking from entering the ionizing region, the filament should be located in a region where the mean pressure is considerably lower than p_2, i.e., close to a pumping lead. Another pumping lead should be connected directly to the tube proper to keep p_s to a minimum value.

The formation of molecular beams has been

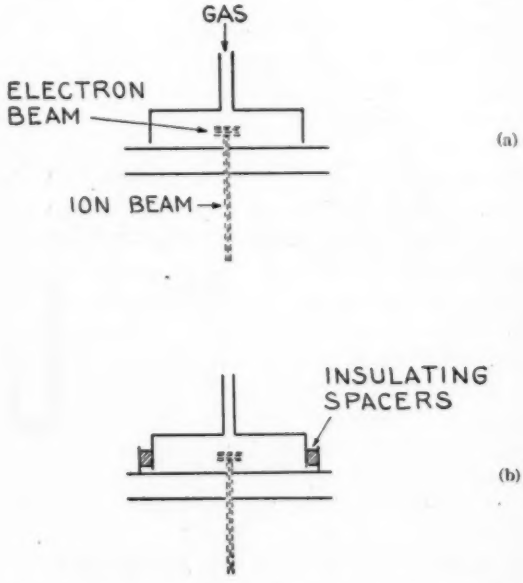

FIG. 7. Geometry of ionizing region. The narrow gap required for electrical insulation in (a) has been eliminated in (b) by the use of insulating spacers.

discussed in detail by Fraser[13] and reviewed by Loeb.[14] It is found that the net rate of molecular flow from region a to b through a tube (Fig. 8) is given as

$$\dot{N}=\tfrac{1}{4}(n_a-n_b)\bar{c}PA, \qquad (11)$$

where

$\dot{N}$ = total number of molecules per sec.,
$n_a=N_a/V_a$ = average molecular density in region a in per cm³,
$n_b=N_b/V_b$ = average molecular density in region b in per cm³,
$\bar{c}$ = average molecular velocity in cm/sec.,
A = area of tube, and
P = probability that any molecule arriving at tube will pass through it.

P becomes unity in the limiting case where the tube is reduced to a hole in a thin diaphragm, and it approaches $4D/3L$ for long tubes of length L and diameter D. Equation (11) assumes that the condition for molecular flow is fulfilled. The spacial distribution of the emerging molecules may be written as

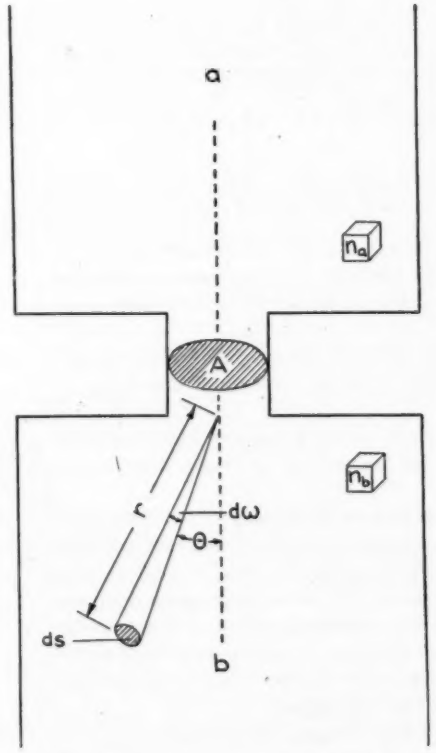

FIG. 8. Molecular beam geometry.

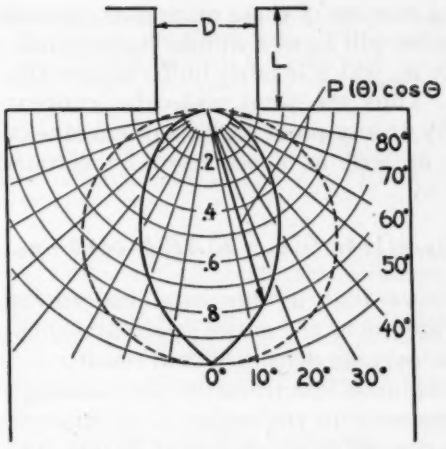

FIG. 9. Plot of $P(\theta)\cos\theta$ as a function of angle θ. The solid curve applies to the special case of a short tube ($L=D$), the dotted circle to a thin diaphragm.

$$d\dot{N}_\theta=\tfrac{1}{4}(n_a-n_b)\bar{c}AP(\theta)\cos\theta\, d\omega, \qquad (12)$$

where

$d\dot{N}_\theta$ = number of molecules emerging per sec. through solid cone $d\omega$ (see Fig. 8),
θ = angle made by $d\omega$ with normal to A, and
$P(\theta)$ = probability for molecules to emerge at angle θ; equal to unity for hole in diaphragm.

Clausing[15] has computed and plotted the quantity $P(\theta)\cos\theta$ in polar coordinates for the special case of a short tube ($L=D$) (see Fig. 9) and compared it to the cosine law valid for diaphragms. While the total gas flow $\dot{N}$ through the short tube, represented by the area under the solid curve, reduces to about one-half of that through the diaphragm, the number of molecules emitted in a forward direction ($d\dot{N}_0$) is undiminished. Thus it becomes clear that well-collimated molecular beams can be readily formed with the help of relatively short tubes.

Equation (12) is easily modified to yield an expression for the molecular concentration $n(r,\theta)$ existing at any point (r,θ) of region b due to the molecular beam. Since $d\omega=ds/r^2$ (see Fig. 8), one may write

$$n(r,\theta)=dN/dV=d\dot{N}(r,\theta)/\bar{c}ds$$
$$=(1/4\pi r^2)(n_a-n_b)AP(\theta)\cos\theta. \qquad (13)$$

If region b is totally enclosed, the molecules in the beam will be scattered by the walls according to the cosine law, as first shown by Knudsen.[16]

[13] R. G. J. Fraser, *Molecular Rays* (The Macmillan Company, New York, 1931), pp. 10–17.
[14] L. B. Loeb, *The Kinetic Theory of Gases* (McGraw-Hill Book Company, Inc., New York, 1934), Sec. 79, pp. 301–310.
[15] P. Clausing, "Über die Strahlformung bei der Molekularströmung," Zeits. f. Physik 66, 471–476 (1930).
[16] M. Knudsen, "Das Cosinusgesetz in der kinetischen Gastheorie," Ann. d. Physik [IV] 48, 1113–1121 (1915).

After a number of these reflections, the scattered molecules will form a diffuse background, of intensity n_0, which is fairly uniform over the entire space. Thus the total molecular concentration $n(r, \theta)_T$ at any point (r, θ) is due to the original beam as well as a scattered background and given by

$$n(r, \theta)_T = (1/4\pi r^2)(n_a - n_b)AP(\theta) \cos \theta + n_0. \quad (14)$$

It follows that if this total concentration is integrated over the entire space and divided by V_b, the average density n_b will result.

In the mass spectrometer, the ionizing region 2 corresponds to the region b, mentioned above in the discussion on molecular beams. The relative importance of the first and second terms of Eq. (14) will depend on the individual spectrometer design. Since part of the molecular concentration, and therefore of the local pressure p_2, depends inversely on r^2, it is obviously desirable to bring the molecular beam as close to the electron beam boundaries as is possible without introducing series distortion of the electric field. This is most readily done by introducing the beam through the top plate, as already indicated in Fig. 7. Since the ion slits in most instruments are quite long, it seems advantageous to replace the cylindrical inlet tube by a tube whose cross section resembles the ion slit. With this geometry, the density of ions produced will be quite uniform over the entire slit length.

It is a pleasure to thank Professor Clark Goodman for reading the manuscript, and making many constructive suggestions. Stimulating discussions with Dr. C. W. Sheppard helped greatly in clarifying several issues. The author also wishes to take this opportunity to thank Dr. J. A. Hipple, Jr., Westinghouse Research Laboratories, the late Dr. O. L. Roberts, Atlantic Refining Company, Philadelphia, and Dr. H. W. Washburn, Consolidated Engineering Corporation, for their cooperation during the construction of the mass spectrometer.

"IF the voice of the people is heard we shall win this peace, but it will be a tougher, longer job than winning the war. For war, after all, is a simple business. Its objectives are clear and heartily supported by the overwhelming majority of citizens. We are united by an impressive common purpose.

"Not so is the struggle to maintain a stable, happy society. Here the objective, while clear enough, has not the dramatic intensity, the immediacy, the urgency to specific individual and collective action. . . .

"Science will march on against disease, ignorance and want, giving us an ever higher material standard of living, but along with scientific thinking we need 'an upsurge of sound, constructive, humanistic thought.'

"The supreme danger lies in the possibility that a group of men, despairing of the competence of their own thinking to meet a particular set of problems and becoming frustrated, will seek cowardly escape of individual responsibility by delegating the authority to think for them to a leader or a central planning authority whom they permit, without safeguards, to control their action and conduct. This voluntary slavery of the mind is more degrading, more inherently vicious and potentially destructive than mere slavery of the person. Its beginnings are insidious. I sense some of them in the organization of various pressure groups, in some of which the leaders are permitted to use methods no self-respecting member of the group would countenance as an individual."

JOHN T. TATE—*Commencement Address*
Case School of Applied Science—April 22, 1945

Here and There

New Appointments

Dr. Lauchlin M. Currie has been elected vice president in charge of research of National Carbon Company, Inc. Dr. Currie has been Acting Director of Research since 1942, except for 15 months during which he was Associate Director of the Division of War Research of Columbia University.

Dr. Earl K. Fischer has begun research work as head of the Division of Physical Chemistry at the Institute of Textile Technology, Charlottesville, Virginia.

Louis J. Chatten, who resigned July 31 as Director of the Radio and Radar Division of the War Production Board in Washington, has been appointed vice president and general commercial manager of North American Philips Company, Inc.

Dr. M. L. Redman has been placed in charge of the Electron Microscope field engineering group of Radio Corporation of America, with headquarters at Camden, New Jersey. For the past two years Dr. Redman has been engaged in supervising RCA Electron Microscope installations in various hospitals and industrial research laboratories throughout the country.

W. James Lyons, secretary of the Division of High Polymer Physics of the American Physical Society, has joined the staff of the Research Laboratory, Firestore Tire and Rubber Company, Akron 17, Ohio. He was formerly with the Radio Corporation of America at Camden, New Jersey.

Gerald Pickett, formerly research physicist with the Portland Cement Association, has taken a position with Kansas State College, Manhattan, Kansas.

James G. Potter, recently a physicist with the Bell Telephone Laboratories, has been appointed head of the department of physics of the Agricultural and Mechanical College of Texas at College Station, effective September 1.

Albert Ray Olpin, executive director of the research foundation of the Ohio State University, has been elected president of the University of Utah. He will take office on July 1, 1946.

W. S. Gillam, professor of agricultural chemistry at Purdue University, Lafayette, Indiana, has joined the staff of the Midwest Research Institute, Kansas City, Missouri, where he will specialize in research in soils, plant nutrition, and analytical services.

John F. G. Hicks, who had been on leave of absence from Corning Glass Works to assist in the development of the atomic bomb, left in August for São Paulo, Brazil, where he will be technical and scientific consultant for Cia. Vidraria Santa Marina, affiliate of Corning Glass Works of South America.

Cyril S. Smith has resigned from his position as research metallurgist for the American Brass Company, Waterbury, Connecticut, to accept appointment as professor of metallurgy and head of the newly formed Institute for the Study of Metals at the University of Chicago. He has been on leave of absence from American Brass for war work since early in 1942, at first as Research Supervisor for the War Metallurgy Committee of the National Research Council, and since April, 1943, with the Los Alamos atomic bomb project near Sante Fe, New Mexico.

Paul Cramer, assistant professor of physics at Denison University, Ohio, has been appointed chairman of the department of mathematics and physics of Huron College, South Dakota.

Howard F. MacMillin has become associated with Arthur D. Little, Inc., Cambridge, Massachusetts. Formerly president and general manager of the Hydraulic Press Manufacturing Company, Mt. Gilead, Ohio, Mr. MacMillin will direct the application to industry of developments in the mechanical engineering and applied physics fields by the laboratories of Arthur D. Little, Inc.

Institute of Textile Technology

The Institute of Textile Technology, a graduate school chartered by the State of Virginia, has announced a building program. The Institute will be enlarged and developed on its present site 1½ miles west of Charlottesville. It is a non-profit corporation, owned by its members and governed by a board of trustees selected by the members, consisting of executives prominent in the textile industry. Its field of service now extends from Maine to Alabama.

The Institute will accept as fellows only applicants who have their Bachelor's degrees with excellent standing from recognized colleges and universities. It contemplates a course of two calendar years leading to a Master's degree and four years leading to a Doctor's degree in the various fields of science. Its prime functions are to train men for the textile industry and to utilize its staff in research in subjects of interest to the textile industry.

New Institutes at University of Chicago

Research which led to invention of the atomic bomb will be continued at the University of Chicago through two new institutes, devoted to the study of nuclear physics and metals. The Institute of Nuclear Studies will be headed by Samuel K. Allison, professor of physics at the University of Chicago, and his staff will include Enrico Fermi, professor of physics, and Harold C. Urey, professor of chemistry. The staff of the Institute of Metals will consist of Cyril Smith, director, and Clarence Zener, professor of metallurgy.

In announcing the establishment of these institutes, Chancellor Robert Hutchins emphasized that the purpose of them is to advance knowledge and not primarily to develop the military or industrial applications of nuclear research. The University will accept contributions from the government and industry for the support of work, but plans to finance the central core of the institutes from its own funds.

As a further outgrowth of its work on the atomic bomb, the University has established an institute for applying the results of nuclear physics research to such problems as cancer, heredity, and the aging process. To be known as the Institute of Radiobiology and Biophysics, this organization will seek also to perfect techniques for the protection of workers using radioactive materials. The Institute will be headed by Professor Raymond E. Zirkle.

Honorary Degree

The degree of Doctor of Science was conferred on Dr. John R. Dunning, professor of physics at Columbia University, at the 56th commencement exercises of the Nebraska Wesleyan University. Dr. Dunning, an alumnus of the university, gave the commencement address.

Administrative Changes at Caltech

Dr. Robert A. Millikan has retired as administrative head of the California Institute of Technology, at the age of 77. He will continue as vice president of the board of trustees and will assist the president of the board in public relations work and institutional development. He plans to devote much time to his own research and writing.

Science reports that Dr. Max Mason has retired from the Executive Council at California Institute of Technology. He will continue as chairman of the Observatory Council and expects to resume work on the 200-inch telescope for the Palomer Observatory.

E. C. Watson, professor of physics, California Institute of Technology, has been appointed dean of the faculty. Other appointments include William V. Houston, chairman, division of physics, mathematics, and electrical engineering; and Frederick C. Lindvall, chairman, division of civil and mechanical engineering.

Death of Eric Loewenstein

Dr. Erich Loewenstein, 58, biological chemist at Bausch & Lomb Optical Company, died at Rochester, New York, following six weeks illness. He was engaged in major research on colorimetric methods of determining vitamin concentration, and had published a number of papers on the subject.

Physical Society of Pittsburgh

Elected officers of the Physical Society of Pittsburgh for the year beginning July 1, 1945, were: Dr. Mary E. Warga, University of Pittsburgh, President; Dr. I. Estermann, Carnegie Institute of Technology, Vice President; Dr. R. C. Mason, Westinghouse Research Laboratories, Secretary-Treasurer.

Physics Club of Chicago

The following are the new officers of The Physics Club of Chicago: President, L. I. Bockstahler; Vice President, W. W. Sloane; Secretary, H. A. Leedy; Treasurer, Louis Robertson.

Technical Translations

The R. T. P.'s listed below are available from the Durand Reprinting Committee in care of California Institute of Technology, Pasadena 4, California.

2428 Wolff, E.	The Effect of Rounding-Off on Originally Sharp Leading Edge on the Resistance of Wings. (Ingenieur-Arch. Vol. IV, No. 6, December, 1933, Part I, p. 521, Part III, pp. 533–537.) (10 pages.)
2429 Sokolik	Detonation and Peroxides in the Internal Combustion Engine. (Acta Physicochemica, USSR Vol. 9, Nos. 3–4, 1938, pp. 593–619.) (15 pages.)
2432 Colombi, Ch.	New Curves of Heat-Energy Transformation. (Bull.-tech. Suisse Romande Vol. 66, No. 14, 13.7.40, pp. 149–154.) (8 pages.)
2433 Gortler, H.	On an Oscillation Phenomenon in Fluids with Stable Density Stratification. (ZAMM Vol. 23, No. 2, April, 1943, pp. 65–71.) (10 pages.)
2439 Gorvisy, A.	The Interaction of Nitrogen with Metals at High Temperatures. (Soc. Chim. de Paris 4th Series, Vol. X, pp. 159–160.) (1 page.)
2441 CZ.	The Electrolytic Welding of Metals. (Gas and Elektrowarme 1944, Pt. 4, pp. 74–75.) (2 pages.)
2443 de Haller, P.	Photographic Striation Methods Applied to the Supersonic Wind Tunnel of the E.T.H. Zurich. (E.T.H. Proceedings No. 8, 1943, pp. 44–49.) (4 pages.)
2444	Optical Multi-Curve Recorder of the Finnish State Aircraft Factory. (Official File.) (7 pages.)
2451 Colombi, Ch.	New Curves of Heat-Energy Transformation. (Bull-tech. Suisse Romande Vol. 66, No. 16, 10.8.40, pp. 169–174.) (8 pages.)
2453 Sokoloff, S. J.	Practical Applications of the Diffraction of Light on Supersonic Waves. (Physikalische Zeitschrift Vol. 36, 1935, pp. 142–144.) (4 pages.)
2454	Regulations for Altitude Flying and Instructions on the Use of Respirators. (D. Luft 1205.) (24 pages.)
2455	The "S.O.30N" and its Design Features. (Aviation Francaise No. 7, 21st March, 1945.) (3 pages.)
2456 Gimm, A.	Fundamental Problems of Aeroengine Design. (Luftwissen Vol. 9, No. 12, 1942, pp. 356–360.) (10 pages.)
2459 Schrader, A.	Structure Investigation in Aluminum Metallurgy. (Aluminum Vol. 25, No. 3, March, 1943, pp. 116–126.) (16 pages.)
2465 Dirksen, B.	On the Development of Play in Bolted Joints under Fatigue. (L.F.F. Vol. 19, No. 4, 6.5.42, pp. 153–156.) (7 pages.)
2466 St. Arnaud, A.	The S.O.30N. A very Special Prototype. (Aviation Francaise No. 5, 7th March, 1945.) (2 pages.)
2467 Eweis, M.	Friction and Leakage Losses in Piston Rings. (Forschung Vol. 6, No. 2, p. 104.) (1 page.)
2468 Hohenemser, K. Prager, W.	The Mechanics of the Plastic Deformation of Mild Steel (ZAMM Vol. 12, No. 1, February, 1932, pp. 1–14.) (17 pages.)
2470 Zeman	The Recent Development of Two-Stroke Engines. (Z.V.D. 1. Vol. 87, No. 1/2, 9.1.43, pp. 7–14.) (12 pages.)
2471 Lubek, H.	The Principles of the Magnetophone Method. (AEC Mitteilungen, No. 9, Sept., 1938, pp. 453–459.) (11 pages.)
2472 Schleicher, F.	Buckling Stresses on Rectangular Fixed Plates. (Mitt. Forschungshutte Outcheftungshutte Konzerns, Vol. 1, 1931, pp. 186–193.) (14 pages.)

New Booklets

The Electronic Corporation of America, 45 West 18th Street, New York 11, New York, has published an illustrated booklet, *The Amazing Electron*, which is available to educational leaders, teachers, and students. Illustrated in cartoon technique and simply written, the booklet makes clear the composition of the atom, the historical development of electronic science, and explains its application in radio (including f-m), television, medicine, and industry. The booklet can be obtained without charge to individuals or for a 5-cent handling charge in quantities of 20 copies or more.

The Taber Instrument Corporation of North Tonawanda, New York, in their Bulletin 4506, announce their new triple cut specimen shear designed for use with the Taber V-5 stiffness gauge to assure accurate cutting of test specimens for uniform and comparable test results.

The Applied Research Laboratories of Glendale, California, and Harry W. Dietert Company, 9330 Roselawn, Detroit 4, Michigan, have published a 4-page folder describing a new two-meter grating spectrograph. This new instrument can be supplied with 36,600 lines per inch or 91,500 total grating lines. Copies available free from Harry W. Dietert Company.

The *General Radio Experimenter* for July and August, 1945, contains articles on "A Heterodyne Frequency Meter for 10 to 3000 Megacycles" and "How Humidity Affects Insulation." 12 pages. General Radio Company, 275 Massachusetts Avenue, Cambridge 39, Massachusetts.

North American Philips Company, Inc., 100 East 42nd Street, New York 17, New York, has published a 12-page booklet entitled *Engineering-Design Development of X-Ray Spectrometer*. The text covers in detail the basic design principles involved in the Geiger-Counter x-ray spectrometer, recently developed industrial control tool. Illustrated with drawings and photos.

Contributed Original Research

The Use of Direct Current in Induction Electron Accelerators

W. F. WESTENDORP
Research Laboratory, General Electric Company, Schenectady, New York
(Received September 14, 1945)

The application of a direct-current bias in the magnetic circuit of induction accelerators is described. In its simplest form this bias reduces power consumption to approximately 30 percent; in especially designed machines a reduction of the weight to 50 percent or less is an additional advantage. A machine with a closed central core is discussed and the use of orbit contraction by saturation over a range of x-ray energies is described.

INTRODUCTION

THE induction electron accelerators or betatrons so far described in the literature[1–7] have been machines operating entirely on alternating current. During one-quarter cycle of the current wave, electrons in a toroidal vacuum tube are accelerated by the electrical gradient associated with the change of magnetic flux and are retained in a circular orbit by the magnetic field between two pole pieces. The machine is utilized only one-quarter of the time even though the electrical gradient is present half the time in the proper direction. Much can be gained in this respect by the use of direct current. Various systems will be described utilizing direct current with a saving of electrical power of more than 50 percent or a saving of materials of well over 50 percent or both. They are described in the order of their electrical and mechanical complication.

I. DIRECT CURRENT ADDED TO THE CONVENTIONAL MACHINE

The superposition of direct current upon the alternating current used in the magnetizing coils of the conventional machine allows the reduction of the alternating current to 54 percent of its original value. It was found experimentally that the iron losses are very nearly proportional to the square of the flux density and that the d.c. component of the magnetic flux did not produce an increase in the hysteresis losses for a given a.c. excitation. This is of importance because the hysteresis losses form the major portion of the losses in the silicon steel used for the magnetic circuit. The a.c. power consumption can therefore be approximately reduced to 30 percent and even with the additional d.c. power requirement the total power consumption will remain below 50 percent.

Figure 1 shows a cross section of the conventional machine. The magnetizing coils form the principal electrical circuit elements. The so-called 2:1 relation* necessary for a circular orbit is established by properly proportioning the gaps between the pole pieces and centerpieces with respect to the main gap in which the vacuum tube is located. The flux density near the orbit follows the equation $B=B_0\rho^{-n}$ in which $\rho=R/R_0$ and $1>n>0$ for stability of the electron orbit or the magnetic focusing of the electrons toward the stable circular orbit.[3]

None of these fundamental requirements interferes with the superposition of direct current

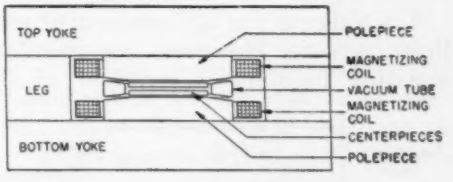

FIG. 1. Cross section of conventional design of betatron in which direct current can be used with a saving of power.

[1] M. Steenbeck, U.S. Patent 2,103,303, Dec. 28, 1937.
[2] D. W. Kerst, Phys. Rev. **60**, 47–53 (1941).
[3] D. W. Kerst and R. Serber, Phys. Rev. **60**, 53–58 (1941).
[4] D. W. Kerst, Rev. Sci. Inst. **13**, 387–394 (1942).
[5] W. F. Westendorp and E. E. Charlton, J. App. Phys. **16**, 581 (1945).
[6] D. Iwanenko and I. Pomeranchuk, Phys. Rev. **65**, 343 (1944).
[7] John P. Blewett, paper to appear in Phys. Rev.

* $2\pi R_0^2 B_0=\Delta\Phi_0$ in which R_0 is the radius of the electron orbit, B_0 the flux density at the circular orbit, and $\Delta\Phi_0$ the change in the total flux through the circular orbit, from the instant at which $B_0=0$.

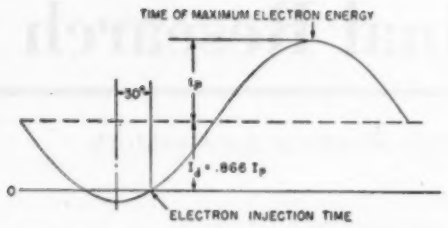

FIG. 2. Magnetizing current as a function of time.

and the corresponding reduction of the alternating component. However, the d.c. component should never exceed the a.c. crest value since otherwise B_θ would not go through zero and no electron injection and capture could take place. Furthermore, it is important to have a considerable electrical gradient present during injection and for this reason it is advisable to make the direct current, e.g., 0.866 of the a.c. crest value with an electrical gradient during injection equal to one-half of the crest gradient occurring during the cycle and locating the injection point 30° beyond the negative crest of the a.c. wave as illustrated in Fig. 2.

Figure 3 shows the electrical hook-up which will allow the superposition of the direct current. The additional circuit elements are

1. a reactor;
2. a d.c. generator or any other source of d.c.;
3. a by-pass capacitor;
4. additional capacitors in the capacitor bank to neutralize the lagging current drawn by the reactor.

Contrary to the first impression the reactor is not of the same order of cost and weight as the machine itself. It can very economically be built as an air-core type and can be either forced air-cooled or convection oil-cooled. The weight of the copper is a function of the R.M.S. value of the total current squared multiplied by the inductance, *viz.* $(I^2L)^{3/5}$, and can be proved to be a minimum for such an inductance as will make the R.M.S. value of the a.c. component equal to the direct current. However, considerably less compensating capacity is required in the capacitor bank for a slightly larger reactor in which the alternating current equals half the direct current. Under these conditions an economic minimum is obtained for the total cost.

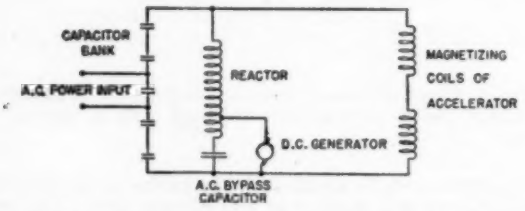

FIG. 3. Circuit diagram for the application of direct current to a conventional machine.

The d.c. source can be a motor generator set or a high current copper oxide rectifier such as used in the electrochemical industry.

The by-pass capacitor need not be very large since its reactive a.c. drop can be 100 percent neutralized by a small fraction of the reactor drop obtained through an appropriate tap, or by means of a few turns around the outside of the non-metallic reactor tank or housing.

II. ADDITIONAL COILS AND SPECIAL POLE PIECES

An analysis of the system first described shows only a saving in power, no saving in materials, and also leads to the conclusion that the maximum electron energy attainable in a given machine is limited by the saturation of the centerpiece laminations which upsets the 2:1 relation and causes orbit contraction. The maximum attainable electron energy has not been changed by the use of direct current but the saturation of the centerpiece is now caused by the combined current a.c.+d.c. If the d.c. component in the central flux could be suppressed and the d.c. component at and near the orbit be retained still higher flux densities could be used at the orbit before causing saturation in the center. This purpose can be achieved by means of direct-current bucking coils located in grooves in the pole pieces as illustrated in Fig. 4. Measurements on a solid steel model with exploring coils and a flux meter indicated the necessity of the use of one large centerpiece and two air gaps. Again the 2:1 relationship is maintained by the proper proportioning of these two air gaps with respect to the main gap. If one gap was used in the center of the model the measurements clearly indicated a horizontal magnetic field component above and below the

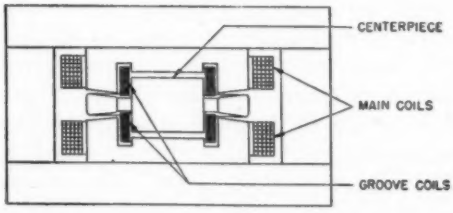

FIG. 4. Machine with grooved pole pieces, extra coils and a large centerpiece.

central plane near the electron orbit caused by the strong direct current in the bucking coils in the grooves. These radial magnetic fields would interfere with the injection and capture of electrons. The construction with the large center-piece and the two gaps eliminated these stray fields since outer pole pieces and center-piece are at the same magnetic potential when the main magnetizing coils carry zero current. Consequently no fringing fluxes would pass through the vacuum tube during electron injection.

Figure 5 shows the electrical connections. The groove coils carry slightly fewer direct-current ampere turns than the main coils since they just buck out the d.c. field in the central core. This requires the same number of ampere turns as are expended in the main gap, whereas the main coils also serve to magnetize the yokes, legs, and the small gaps between the various parts of the magnetic circuit.

The additional parts are an a.c. blocking filter consisting of a reactor and a capacitor, a d.c. generator, and the groove coils. The pole-piece construction with grooves offers an additional mechanical problem. However, it can be shown that the weight of the silicon steel for, e.g., a 50-Mv machine can be reduced to approximately 50 percent of the weight of the a.c. design, while for higher energies the reduction is even greater.

The blocking filter can economically be designed so that the a.c. current through the reactor is 50 percent of the direct current. The exciting current of the filter can readily be made less than 0.5 percent of the direct current. The exciting current will flow through the d.c. generator and through the groove coils. It will have a negligible effect on the location of the orbit.

Figure 6 shows a different circuit which permits all the power capacitors to be placed in one bank. In this case the groove coils have been given a higher number of turns than the main coils so that the induced e.m.f. in the groove coils equals the voltage across the main coils and consequently no alternating current will flow through the groove coils. However, in order to keep the d.c. ampere turns at the proper value, the groove coils now will have to carry a lower direct current. A control resistance and a by-pass reactor allow adjustment of the direct current through the groove coils. It is possible to design a machine in which the reactor and the resistor can be omitted; however the d.c. ampere turns in the groove coils would over-compensate the d.c. ampere turns in the main coils and a d.c. flux through the centerpiece would result, causing saturation during the negative peak of the a.c. current wave. The magnetic circuit of such a machine without reactor would be considerably larger than the one for which the complete diagram of either Fig. 5 or Fig. 6 is carried out.

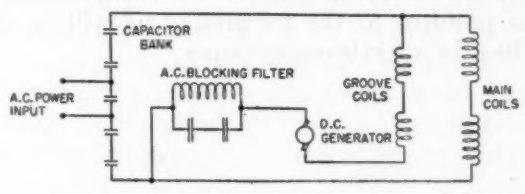

FIG. 5. Circuit diagram for the machine of Fig. 4.

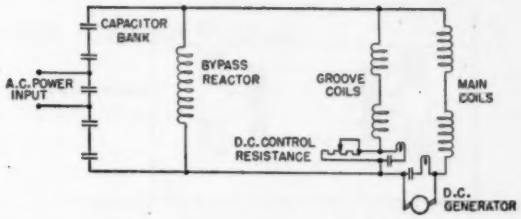

FIG. 6. Circuit diagram applicable to the machines of Figs. 4 and 7.

In Fig. 6 the conductor loops shown in series with the by-pass capacitors across the d.c. control resistance and the d.c. generator represent turns coupled with the reactor in order to compensate the capacitive voltage drop.

III. ELIMINATION OF THE CENTRAL GAPS

Examination of Fig. 6 leads to an interesting consequence of this diagram. The groove coils have a larger number of turns than the main coils in order to match voltages and are directly connected across the main coils as far as the a.c. operation is concerned. The 2:1 relation now depends no longer on the ratio of magnetic reluctances but is fixed by the turn ratio of the groove and main coils, their voltages being identical and their magnetic fluxes being in inverse proportion to their numbers of turns. A fine adjustment of the electron orbit diameter as determined by the correct 2:1 relation may be made by adjusting the flux ratio by means of an additional tapped autotransformer connected between main coils and groove coils. This autotransformer is not shown in Fig. 6.

Figure 7 shows a cross section of a machine in which the central gaps are eliminated and Fig. 8 shows a model. For ease of manufacture the two pole pieces and the centerpiece are still three

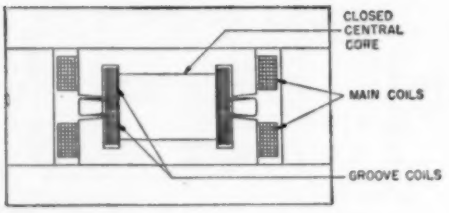

FIG. 7. Machine with closed central core.

separate units but the gaps are reduced to the lowest practical minimum so that a sheet of kraft paper or thin pressboard can be used as a separator. The groove coils in this machine will have to carry both d.c. and a.c. bucking ampere turns since the reluctance of the closed central core requires very few ampere turns for magnetization and therefore an alternating current will flow counteracting the effect of the main coils. More space is therefore needed for the groove coils. However, the elimination of the central gaps results in a reduction of the magnetic energy stored in the machine at the peak of the current wave. This energy is stored principally in the air gaps where the product HB is high. By the elimination of the central gaps this stored energy can be reduced by 50 percent or more depending on the space taken up by the vacuum tube. The final result will be an appreciable reduction in the size of the capacitor bank.

FIG. 8. Solid steel model of a betatron with ring-shaped pole piece, groove with coil and large centerpiece. Top yoke and top pole piece have been removed to show details. This model was used to study the field distribution with direct current excitation, by means of a pick-up coil and a flux meter.

The use of the circuit diagram of Fig. 6 in conjunction with the machine shown in Fig. 7 has the added advantage of adjustment of the electron and x-ray energy when orbit contraction by means of saturation of the central core is used.[2] A relatively small reduction in the direct current in the groove coil by means of the control resistor will cause saturation of the central core at the crest of the positive half cycle of the a.c. wave since the ampere turns of the main coil are then insufficiently compensated by the groove coil current. The fact, that the direct connection of groove coils to main coils tends to maintain the 2:1 relation, will cause a distortion in the a.c. wave form of the current drawn by the groove coils. This distortion is limited by the relatively high leakage reactance of this winding and does not offset the effect of saturation.

It is possible to produce saturation of the central core over a wide range of values of the main a.c.+d.c. operating current. The ratio of a.c. to d.c. crest is maintained throughout this range at, e.g., 0.866 and the control resistor in the groove coil circuit is used to produce saturation at the crest of the wave for each setting of the main current. A wide range of output x-ray energies can thus be covered in a simple manner without resort to orbit shift circuits involving high current vacuum tubes.[4,5]

CONCLUSION

Various methods have been described which will make possible the construction of induction accelerators with peak flux densities at the orbit of 6000 to 8000 gausses and correspondingly smaller radii resulting in considerable saving of materials and electric power.

For machines of more than 50-Mv electron energy such a reduction in the size of the orbit will introduce the new problem of the radiation by high speed electrons in a circular path.[6] Experience with the 100-Mv induction accelerator in our laboratory[5] has clearly brought this problem to the foreground as will be described in a forthcoming paper.[7]

Crucible Failure in the Induction Melting Process

ANDREW GEMANT AND JOSEPH STICHER
The Detroit Edison Company, Detroit, Michigan

(Received June 4, 1945)

Experience shows that in the course of the induction melting process occasional local overheating of the crucible takes place, which in extreme cases leads to perforation of the latter. An explanation of this occurrence is presented on the basis that it is essentially a phenomenon of thermal instability. The theory shows that overheating of the crucible is more probable the steeper the electrical resistivity—temperature characteristic of the crucible material, a condition brought about by impurities present. Numerical computations, showing the transition from stable to unstable operating conditions, are carried out for a small experimental and a large industrial furnace. The equations developed show how the various operating conditions, particularly the frequency of the current in the coil, have to be selected in order to insure operation in the stable region.

1. THE PROBLEM

INDUCTION heating is rapidly developing into a highly important tool in engineering processes. The versatility of the method is shown in a paper by Sherman,[1] and is emphasized also by Baker.[2] A particular application, with which this paper is concerned, is the high frequency induction furnace. Until recently this method was chiefly used for steel alloys, hence up to temperatures of 1600°C, but nowadays it is commonly used for alloys requiring temperatures appreciably beyond that limit. The question of the crucible material is then of considerable importance; zirconium oxide crucibles, with a melting point of around 3000°C, are among those that have been developed for this purpose.

Experience has shown that occasional trouble arises caused by local overheating of the zirconia crucible, which causes the latter to melt and eventually allows the molten metal charge to run out. This local disturbance occurs while the bulk of the charge and of the crucible is appreciably below the melting point of the latter. The present paper is concerned with a discussion and explanation of such crucible failures in induction heating.

In ordinary heating processes, the heat supplied is constant and the temperature of the material being heated rises until the rate of heat dissipation is equal to the rate of heat input or, in other words, until equilibrium conditions are established. The basis for the explanation of the phenomenon discussed in the previous paragraph lies in the electrical conductivity *versus* temperature characteristic of the crucible material. This characteristic may be such that the portion of the heat that is generated in the crucible may not be constant. It may, indeed, as the crucible material becomes more and more conducting with rise in temperature, increase more rapidly than the heat dissipating characteristic. Thus equilibrium conditions cannot be reached and the temperature continues to rise until the crucible material melts.

It should be mentioned that an explanation as here suggested occurred to the authors as an analogy to the so-called thermal breakdown of electrical insulators. This latter mechanism was first quantitatively developed by K. W. Wagner[3] and extended, among others, by Gemant and S. Whitehead.[4,5] In the case of insulation it is an electric field due to high potential electrodes that causes a current to flow, develops heat, and leads to failure; in the case of induction melting it is the electromagnetic field that by a similar process leads to failure.

In the process of induction melting, conditions are generally such that stable operation is possible. However, in localized spots of the crucible conditions might be different, either owing to inhomogeneity of the material or to secondary reactions, for instance, reduction of the zirconia. If in this localized region conditions of stability are not fulfilled, local failure will be the result, even though the bulk temperature of the crucible is below its melting point.

Although in practice the failure is probably always a localized process, the mathematical treatment given herein will be based on homogeneous conditions extended over the whole circumference. The reason for this postulate is that the results probably do not differ essentially in

[1] V. W. Sherman, Ae. Eng. Rev. **2**, No. 5 (1943).
[2] R. M. Baker, Trans. A.I.E.E. **63**, 273 (1944).
[3] K. W. Wagner, J. Am. Inst. Elec. Eng. **41**, 1034 (1922).
[4] K. W. Wagner and A. Gemant, Preuss. Akad. Ber. (1934).
[5] Andrew Gemant and S. Whitehead, Phil. Mag. **27**, 582 (1939).

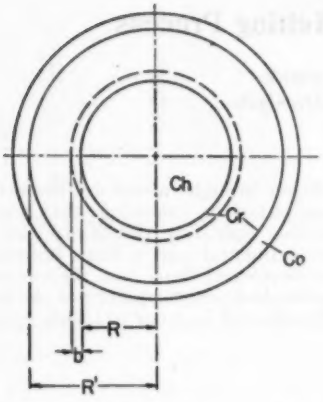

FIG. 1. Cross-sectional diagram of induction heating furnace.

the two cases, and that the simplified treatment as presented will be sufficient to show the principal features of the mechanism. Besides, the treatment remains simple only so long as circular symmetry holds, becoming too involved for simple mathematical analysis when conditions vary along the circumference of the crucible.

This presentation shows in a semi-quantitative manner the conditions that must be fulfilled to make stable operation possible. One of the essential characteristics is the frequency of the applied current, and the results indicate how the conditions of stability vary with the frequency. This circumstance permits the selection of a frequency range assuring optimum operating conditions.

2. THE CRITERION FOR THERMAL STABILITY

The treatment of the problem given refers to a cylindrical charge arranged coaxially with a cylindrical coil (Fig. 1). In the cross-sectional figure, *Ch* indicates the charge, *Co* the coil, and the space *Cr* between them the crucible. The radius of the metal charge is R, that of the inner surface of the coil is R'. The space *Cr* is either completely occupied by the crucible, or the crucible is only the inner part of the space, as indicated by the dotted circle, the space outside the crucible being filled with a granular mass acting as spacer and consisting of ZrO_2 pebbles or the like. The wall thickness of the crucible in the former case is $R'-R$, in the latter b.

In the subsequent analysis it is assumed that the device is sufficiently long for end effects to be negligible in a first approximation. This is more justified, the better the heat insulation at the top and bottom of the furnace. Hence, heat flow is assumed to take place only in a radial direction. All data concerning heat generated and dissipated refer, therefore, to unit axial length (cm).

The consideration of transient conditions is outside the scope of this paper, in other words, the process of the gradual heating of the charge is not considered. The question is only whether or not conditions permit stable equilibrium.

Two quantities will have to be computed, the heat, W_g, generated per second per cm length and the heat, W_d, dissipated per second per cm axial length, in order to determine under what conditions $W_g = W_d$, the criterion of stability. In making use of the theory of induction heating, the reader is referred to various texts, or to articles such as that by Greig.[6] The induction melting process in particular is discussed by Wall.[7]

The quantity W_g is the sum of two terms, the heat generated in the charge, W_{Ch}, and in the crucible material itself, W_{Cr}. The heat generated in the coil does not have to be considered in this connection.

For the value of W_{Ch} the following equation will be used:

$$W_{Ch} = (RH_0^2/8)(f\rho_{Ch})^{\frac{1}{2}}, \qquad (1)$$

where H_0 = amplitude of the magnetic force in the empty coil (in oersteds), f = frequency of applied current in coil, ρ_{Ch} = electrical resistivity of charge (electromagnetic units). This relation is valid only if the condition

$$\pi R(8f/\rho_{Ch})^{\frac{1}{2}} > 4 \qquad (2)$$

holds. Expressed in words, formula (2) states that the ratio of the charge radius to the depth of penetration of the magnetic field should be larger than, say, 4. Conditions of operation are nearly always such that (2) is satisfied, and, accordingly, Eq. (1) can be used.

The permeability in Eq. (1) as in all following equations is taken as unity. This is correct for non-magnetic charges, such as non-ferrous metals and ferrous metals at temperatures above the Curie point (about 800°C).

Concerning W_{Cr}, it has to be pointed out that in general the total temperature drop takes place across the crucible, since the coil in practice is usually cooled to a relatively low temperature. While the electrical resistivity, ρ, of the crucible

[6] James Greig, World Power **25**, 246 (1936).
[7] T. F. Wall, The Engineer **173**, 209 (1942).

is very high at low temperatures, it falls to low values at high temperatures (more will be said about this relation in Section 3). The current flowing in the crucible is, therefore, negligibly small except for a narrow region adjacent to the charge that has a temperature only slightly below that of the metal. A rigorous treatment would have to start with the differential equation of the heat balance in each annulus of the crucible. A good approximation could be carried out on the basis of the logarithmic drop of temperature across the crucible, but the present treatment is even further simplified.

It is assumed that a thin annulus adjoining the charge is the only heat source due to the crucible. This amount of heat, W_{Cr}, is then added to W_{Ch} and the total is considered as a cylindrical heat source acting at the boundary between the charge and the crucible. The bulk of the crucible is considered free of heat sources from the standpoint of heat conduction. In the case of a thin crucible in a granular mass the thin annular source is simply identified with the crucible; the thickness of the source is then b (Fig. 1). In the case of a thick-walled crucible the thickness of the source (also designated by b) is taken arbitrarily. The value of b in this case might also be estimated by inspecting the depth of discoloration that can be observed at the inner wall of crucibles.

Since the resistivity of the crucible, ρ, at any point, even at the inner surface, is appreciably larger than ρ_{Ch}, it is probably correct to say that the magnetic field in the crucible is H_0, the same as in a coil unaffected by the charge. This statement is utilized in computing W_{Cr}.

If i_0 denotes the maximum of current density in the crucible material, then $i_0^2\rho/2$ is the heat generated per cc. As to the current density, i:

$$i=-(d\phi/dt)/2\pi R\rho \tag{3}$$

(with permeability unity), where ϕ is the flux linked with the annulus of radius R and thickness b.

The flux ϕ is the sum of two terms, the flux in the charge, ϕ_{Ch}, and the flux in the thin layer of crucible, ϕ_{Cr}. For the first it is assumed that, because of the large ρ/ρ_{Ch} ratio, its value is not affected by the presence of the crucible; we then have (with t = time):

$$\phi_{Ch}=\frac{H_0R}{2}\left(\frac{\rho_{Ch}}{f}\right)^{\frac{1}{2}}(\cos 2\pi ft+\sin 2\pi ft). \tag{4}$$

This relation is valid if condition (2) holds; otherwise certain Bessel functions have to be considered.

Concerning ϕ_{Cr}, the field intensity in the crucible is $H_0\cos 2\pi ft$. Evidently for the innermost layer of the annulus b this flux component is zero; for the outermost layer it is $2\pi RbH_0\times\cos 2\pi ft$. Taking the arithmetic average:

$$\phi_{Cr}=\pi RbH_0\cos 2\pi ft. \tag{5}$$

The sum of Eqs. (4) and (5) is the total flux linked with the annulus. From Eq. (3),

$$i=\frac{H_0}{2\rho}[-(\rho_{Ch}f)^{\frac{1}{2}}\cos 2\pi ft+((\rho_{Ch}f)^{\frac{1}{2}}+2\pi fb)\sin 2\pi ft]. \tag{6}$$

Adding the two vector components:

$$i_0=\frac{H_0}{2\rho}[\rho_{Ch}f+((\rho_{Ch}f)^{\frac{1}{2}}+2\pi fb)^2]^{\frac{1}{2}}. \tag{7}$$

Computing $i_0^2\rho/2$, the heat generated per cc becomes:

$$H_0^2k/8\rho,$$

with

$$k=\rho_{Ch}f+((\rho_{Ch}f)^{\frac{1}{2}}+2\pi fb)^2. \tag{8}$$

Hence,

$$W_{Cr}=\pi RbH_0^2k/4\rho. \tag{9}$$

The total heat generated, W_g, is then the sum of Eqs. (1) and (9):

$$W_g=\frac{RH_0^2}{4}\left(\frac{\pi bk}{\rho}+\frac{(\rho_{Ch}f)^{\frac{1}{2}}}{2}\right). \tag{10}$$

The next step is the calculation of W_d, the heat dissipated radially across the crucible toward the coil. We have

$$W_d=G(T-T_0), \tag{11}$$

where T = temperature of the charge, T_0 = temperature of the coil, both in absolute degrees, and G = thermal conductance of the crucible. For G we have

$$G=(g/2\pi)\log(R'/R), \tag{12}$$

with g = thermal conductivity of the crucible material.

In the following analysis ρ_{Ch}, the electrical resistivity of the metal, is taken as a constant, its moderate increase with temperature being disregarded. The temperature dependence of ρ on the other hand is considerable, and will be

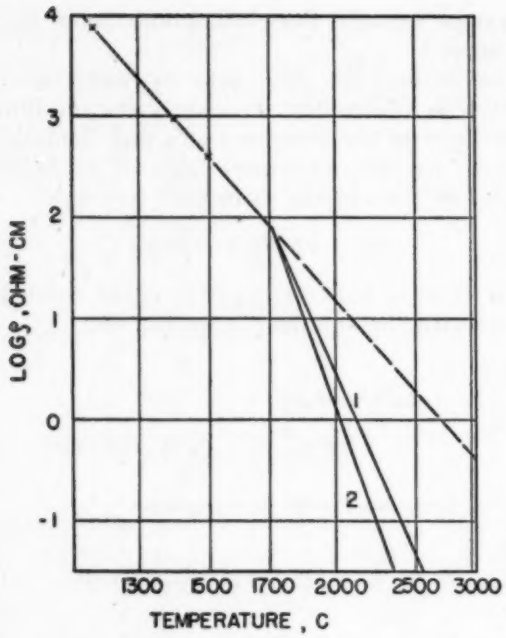

FIG. 2. Specific resistivity of zirconia as a function of temperature. Dotted line: straight extension of experimentally known portion; (1) and (2): assumed breaks, according to information on other inorganic crystals.

represented by the equation

$$\rho = \rho_0 \exp(\beta/T), \tag{13}$$

with ρ_0 and β = constants.

In order to determine the criterion of stability, both W_g and W_d are considered as functions of temperature. If plotted in this manner, one single curve results for W_d, but a whole family of curves for W_g, containing H_0 as the parameter. Most of the W_g curves show two intersections with W_d: the one, corresponding to the lower temperature, represents the stable state, i.e., the equilibrium temperature at that particular value of H_0; the other is of no significance, since it determines an unstable equilibrium. The W_g curve is being shifted toward higher values with increasing H_0. If a certain critical value of the latter which may be called the limiting field H_l is reached, W_g and W_d will just touch at one point. For still higher fields there will be no common points between the two curves, i.e., the temperature rises continuously without ever reaching an equilibrium.

From this analysis the method of finding H_l is evident. It must fullfill two conditions, one being

$$W_g = W_d,$$

the other

$$dW_g/dT = dW_d/dT.$$

These two equations contain two unknowns: H_l and T_l (the limiting temperature) which can be calculated.

The two equations are:

$$\frac{RH_l^2}{4}\left(\frac{\pi bk}{\rho_0 \exp(\beta/T_l)} + \frac{(\rho_{CK} f)^{\frac{1}{2}}}{2}\right) = G(T_l - T_0), \tag{14}$$

and

$$\frac{\pi\beta RH_l^2 bk}{4T_l^2 \rho_0 \exp(\beta/T_l)} = G. \tag{15}$$

Eliminating H_l, one obtains

$$\frac{\rho_0 \exp(\beta/T_l)(\rho_{CK} f)^{\frac{1}{2}}}{2\pi bk} + 1 = \frac{\beta(T_l - T_0)}{T_l^2}, \tag{16}$$

which allows computation of T_l. The limiting field is then from Eq. (15)

$$H_l = \left(\frac{4T_l^2 \rho_0 \exp(\beta/T_l)}{\pi\beta RbkG}\right)^{\frac{1}{2}}. \tag{17}$$

Equation (17) in conjunction with Eqs. (8) and (16) contains the final result of our analysis.

3. THE ELECTRICAL RESISTIVITY OF ZIRCONIA

Of the various quantities occurring in Eq. (17), the frequency of the applied current plays implicitly an essential part, as will become evident from subsequent numerical computations. As explained, b is arbitrary, but this weakness could be remedied by a more elaborate treatment.

The most important constant is perhaps β, introduced by Eq. (13), since it occurs in the exponent. Unfortunately, its value is not known from experiment, and for this reason this subject deserves a brief discussion.

The electrical conductivity of zirconia has been measured only up to 1500°C, as far as the authors could ascertain. The few existing data[8] are plotted in Fig. 2. As pointed out, however, recent applications extend up to 1800°–1900°C, and it is rather unlikely that the same β (slope) that is valid up to 1500° should hold up to higher temperatures (dotted line).

There is information available[9] on a large

[8] International Critical Tables (McGraw-Hill Book Company, Inc., New York, 1927), Vol. II, p. 86.

[9] Landolt-Börnstein, *Physikalisch-Chemische Tabellen* (Verlagsbuchhandlung Julius Springer, Berlin, 1936), Complem. Vol. IIIc, p. 2017.

number of lower melting crystals for which the data cover the whole temperature range up to the melting point. These crystals are essentially the halides of alkali metals, silver, thallium, and the like. It appears that at temperatures of several hundred degrees C below the melting point the slopes of the curves suddenly increase to values of about twice the previous values. Lacking experimental data on ZrO_2 for temperatures above 1500°C, two arbitrary extensions (lines 1 and 2) were assumed as possible bases for numerical calculations. Because of these assumptions the following numerical data can be considered as only semi-quantitative. A research filling this gap in our knowledge would be very valuable.

It is known that admixtures to ZrO_2 lower the resistivity considerably (the values for the Nernst glower—essentially ZrO_2 with a few other oxides—for instance, are 1000 times lower than those for pure zirconia). The actual ρ *vs.* T curve will, then, depend on the degree of purity of the material, the resistivity being lower the more impurities are present. The lines 1 and 2 represent two possible curves corresponding to two different degrees of contamination. Secondary processes, as mentioned in Section 1, result in a lowering of the ρ *vs.* T curve, and this, in its turn, influences the values of ρ_0 and β in Eqs. (13) and (17). It will be shown in the following section how the value of these constants influences the conditions determining stability.

4. NUMERICAL COMPUTATIONS FOR TWO INDUCTION MELTING FURNACES

In the following, two special cases are discussed and the equations deduced above are used in determining the limiting field intensity, beyond which stability is not possible. The ampere-turns (r.m.s. value) per cm axial length, AS, are related with H_0 by the equation

$$AS = H_0/0.4\pi\sqrt{2}, \qquad (18)$$

and the final results will be presented in ampere-turns.

The numerical values for those various constants that occur commonly in both cases are as follows: One of the higher melting metals for which sufficient data are available in the literature is tungsten, and this metal is used, therefore, in these examples. The electrical resistivity of tungsten, ρ_{CA}, at around 2000°C is 60×10^{-6} ohm cm or 60×10^{3} electromagnetic units. The magnitude of b is taken 0.3 cm, this being the order of magnitude of the wall thickness of small size crucibles. The outer surface of the container, adjoining the tubes carrying cooling water, has a temperature of, say, 100°C or, in a rounded number, $T_0=400$°K.

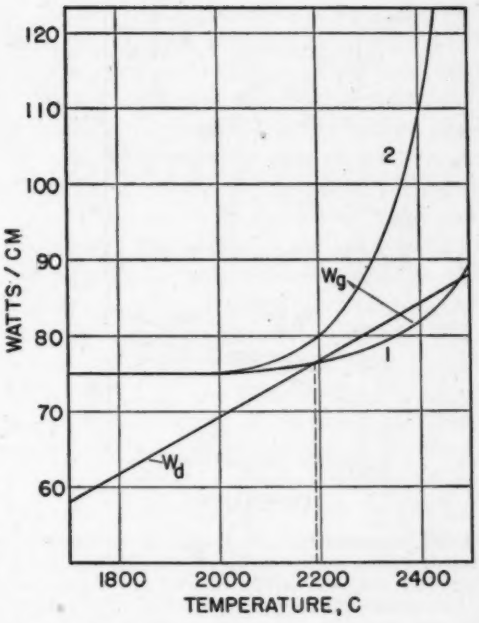

FIG. 3. Heat generated and dissipated *vs.* temperature for a crucible with a radius of 2 cm, a frequency of 4×10^5 c.p.s., and 78 amp.-turns/cm. Curve (1) shows stable, curve (2) unstable operating conditions.

Concerning ρ, the resistivity of ZrO_2, slopes (1) and (2) of Fig. 2 are used, as pointed out above. For curve (1), ρ_0 is 0.32 electromagnetic unit, $\beta=53\times10^3$, for curve (2), ρ_0 is 0.8×10^{-3} electromagnetic unit, $\beta=65\times10^3$.

(*a*) The first case under consideration is an experimental type of furnace, utilizing a small zirconia crucible of 2 cm radius, surrounded by zirconia pebbles in a thickness of 6 cm ($R=2$, $R'=8$). The thermal conductivity, g, of the zirconia pebbles was taken as 0.8×10^5 ergs per sec. per cm per deg. C or 0.5 B.t.u. per hr. per ft. per deg. F, as estimated from data on zirconia brick.[10] From Eq. (12), $G=3.7\times10^5$ ergs per sec. per cm per deg. C.

The results of a detailed calculation are shown in Fig. 3, referring to a frequency of 4×10^5 c.p.s., a high value suited to the small size of the charge. The curve shows watts/cm *vs.* tem-

[10] Max Jakob and G. A. Hawkins, *Elements of Heat Transfer and Insulation* (John Wiley & Sons, Inc., New York, 1942).

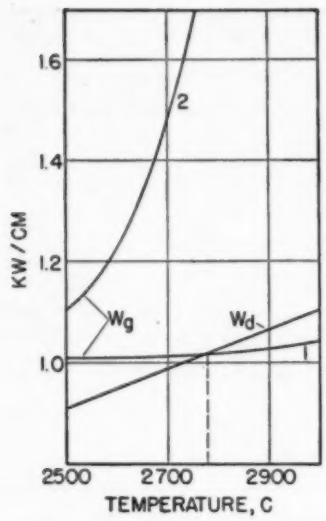

FIG. 4. Heat generated and dissipated *vs.* temperature for a crucible with a radius of 25 cm, a frequency of 4×10^3 c.p.s. and 260 amp.-turns/cm. Curve (1) shows stable, curve (2) unstable operating conditions.

perature in centigrades. The line W_d is the heat dissipated. For W_g branch (1) in Fig. 2 was assumed. Calculating the limiting field and the limiting ampere-turns, AS_l, a value for the latter of 80 was obtained from Eqs. (17) and (18). This means that stable operation is possible only below that value. As an example, $AS=78$ was selected—a value slightly below 80—and the curve (1) as shown was calculated from Eq. (10). The left-hand intersection indicates the condition for equilibrium. The steady-state temperature under these conditions would be somewhat below 2200°C.

Suppose now that at one localized spot the crucible contains impurities that cause the $\rho-T$ characteristic to follow branch (2) of Fig. 2. Calculating W_g for such a case (with $AS=78$), curve (2) of Fig. 3 results, showing that no stable condition is possible. The temperature at that localized spot will rise until the crucible melts and and the charge runs out. Calculating AS_l for this case, a value of 77 is obtained. Hence, with our ampere-turns of 78, it is obvious that we are now in the unstable region.

(*b*) The second case refers to a large, industrial furnace, having a charge radius of 25 cm and a crucible thickness of 8 cm. The thermal conductivity of the solid crucible is taken as 1.7×10^5 ergs per sec. per cm per deg. C or 1.0 B.t.u. per hr. per ft. per deg. F. From Eq. (12), $G=3.8\times10^6$ ergs per sec. per cm per deg. C.

The results for this case are shown in Fig. 4 which is analogous to Fig. 3. The frequency is 4×10^3 c.p.s. With branch (1) of the $\rho-T$ characteristic (Fig. 2), $AS_l=270$ amp.-turns/cm. Thus, we apply only 260 amp.-turns, and obtain curve (1), indicating stable operation. Assuming branch (2) of Fig. 2 to be valid at a localized spot, curve (2) results, having no intersections with the W_d curve. AS_l for this latter case turns out to be 250 amp.-turns/cm.

When the failure point is reached in a furnace of this type, pitting of the inner surface of the crucible, rather than perforation takes place. This process, however, might be cumulative and eventually leads to a destruction of the crucible.

5. PRACTICAL CONCLUSIONS

Practical conclusions from the analysis presented in this paper can be discussed from two angles.

(1) If the explanation of crucible failure, as given, is correct, it follows that the main reason for a failure is the steepness of the resistance—temperature characteristic of the crucible material at temperatures above 1600°C. This again is influenced by the degree of purity of the zirconia. These impurities are either originally present in the material or are introduced as a

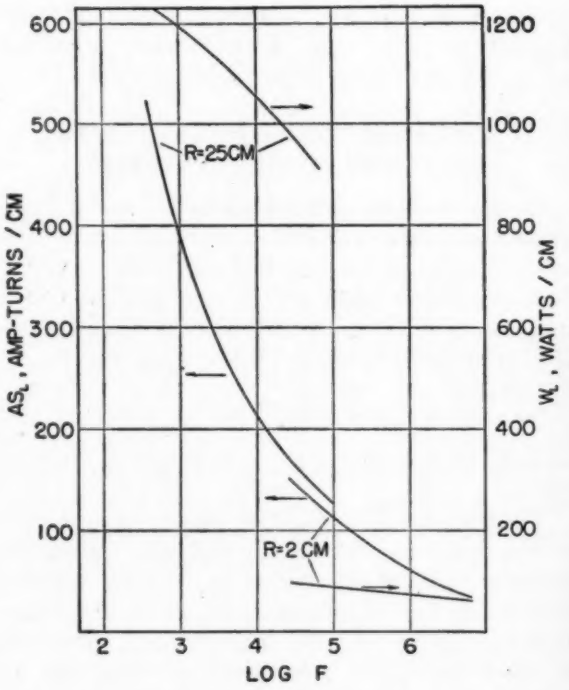

FIG. 5. Limiting amp.-turns/cm and limiting watts/cm *vs.* frequency for crucibles of 2- and 25-cm inner radius.

result of secondary chemical reactions during the melting process.

If troubles of this type are to be avoided, the quality of the zirconia has to be improved by proper selection of the raw materials and manufacturing processes, and operating conditions have to be selected so as to prevent the occurrence of unwanted reactions.

(2) If an improvement along these lines is not feasible or is otherwise unwarranted, the process has to be carried out under conditions permitting a high value of AS_l. The role of the various factors is evident from Eqs. (16) and (17) and a systematic analysis would show the effect of each factor upon the value of AS_l quantitatively.

As an example, the role of the frequency, one of the most essential factors, is shown in Fig. 5, both for the experimental set-up ($R=2$) and the large furnace ($R=25$). The abscissa gives the logarithm of the frequency which is chosen as the independent variable. The ordinates show (left side) the respective values of AS_l, the limiting ampere-turns/cm, and (right side) W_l, the corresponding watts/cm generated in the charge, calculated from Eq. (1). This latter quantity determines the speed of the melting process, the time to reach a given temperature being shorter the larger the heat input rate.

The range of possible frequencies used is limited toward the lower end of the range by conditions imposed by the depth of penetration. These conditions are known and need not be discussed here. The present analysis shows that there is equally a limit toward the high end of the frequency range. The higher f, the lower both the limiting ampere-turns/cm and the limiting heat input W_l. The conclusion can be drawn from Fig. 5 that the lowest frequency compatible with the depth of penetration criterion permits the highest limiting input. If, however, other considerations suggest the use of higher frequencies, as in practice often might be the case, the graph shows how far the ampere-turns can be increased without damage to the crucible.

It is evident from the section dealing with the resistivity of zirconia that the graphs presented in this paper will have to be recalculated when true resistivity—temperature characteristics of the crucible material are obtained experimentally.

A Double-Solenoid Coil System for Magnetic Measurements

E. W. Frowe and C. J. Aronson
Naval Ordnance Laboratory, Navy Yard, Washington, D. C.
(Received August 2, 1945)

A double-solenoid coil system is described which can be used as a neutralizing coil having large internal and small external magnetic fields. Its application in the measurement of magnetic moments of small magnets is also discussed.

INTRODUCTION

1. EXPERIMENTAL techniques occasionally necessitate the use of current-carrying coils to neutralize magnetic fields within very limited volumes. Since it is often not desirable for the effect of the neutralizing coil to extend appreciably beyond the volume where the field is to be neutralized, an ordinary solenoid coil cannot always be used as a neutralizing coil.

2. One of the solutions to the problem of obtaining a neutralizing coil having a limited extent of external field employs a toroid-shaped coil to provide the neutralizing field. In Fig. 1 volume C represents a region within the toroid in which the component of field parallel to the y axis is to be neutralized. The distances A and B are measured from the center of the volume C to the top and side of the toroid, as shown. The volume C is small enough, so that the curvature of the neutralizing field of the toroid may be neglected. The principal disadvantage of such a system is that to produce a uniform neutralizing field in the volume C, it is necessary to have a comparatively large toroid, i.e., $A \gg B$, precluding the use of such a coil within small spaces. A second disadvantage is that the toroid neutralizes only the component of field parallel to the y axis, as shown. If this component is parallel to a surface, it can be neutralized at a distance B from the surface; but if this component is perpendicular to a surface, it cannot be

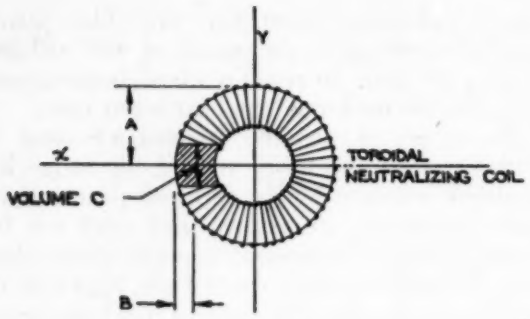

FIG. 1.

neutralized closer than the distance A from the surface. A third disadvantage of the toroid is that such a coil is difficult to wind. The double solenoid was developed to overcome these disadvantages of the toroidal neutralizing coil.

DESCRIPTION OF THE DOUBLE SOLENOID

3. The double-solenoid neutralizing coil which is shown schematically in Fig. 2 consists of two coaxial, solenoidal windings of the same length, L, but with different diameters and different numbers of turns. The two solenoids are connected in series, so that they produce magnetic fields of opposite polarity. The design is such that the magnetic moment (product of the total number of turns times the cross-sectional area) of one solenoid is equal to the magnetic moment of the other solenoid. Thus, the external field of the system decreases rapidly with distance from the center and becomes negligible for most purposes even at distances as small as one solenoid length from the center of the coils. The internal field, however, is not zero, but is equal to the difference in the fields of the two coils. The inner coil, having the greater number of turns per centimeter and the smaller radius, produces the stronger field at the center of the system.

FIELD AND MOMENT CALCULATIONS

4. Computations of the internal field of the double-solenoid neutralizing coil are based on the equation for the field, H, at the center of a solenoid of finite length, L centimeters,

$$H=0.4\pi NI\cos\alpha, \tag{4.1}$$

where H=magnetic field intensity, in oersteds, at the center of the solenoid, N=number of turns per centimeter of length, I=current, in amperes, and α=the angle determined by the axis of the solenoid and a line intersecting the axis at the center of the solenoid and passing through the winding at the end of the solenoid (tangent α=the radius divided by the half-length of the solenoid). The magnetic moment M in e.m.u. of a solenoid is given by the relation

$$M=0.1nIa, \tag{4.2}$$

where n=total number of turns=NL, a=cross-sectional area of the solenoid in square centimeters (πR^2 for solenoid of radius R centimeters), and I=current, in amperes. Figure 2 shows a double solenoid of length L, having an inner winding of radius R_1 and N_1 turns per centimeter, and an outer winding of radius R_2 and N_2 turns per centimeter. From relation (4.1) and the geometry of Fig. 2 it can be shown that the net field, H, of a double solenoid is equal to the difference between the fields of the inner solenoid, H_1, and the outer solenoid, H_2,

$$H=H_1-H_2=0.4\pi IL[N_1/(L^2+4R_1^2)^{\frac{1}{2}} - N_2/(L^2+4R_2^2)^{\frac{1}{2}}]. \tag{4.3}$$

From relationship (4.2) and the geometry of Fig. 2, it may be shown that the net moment, M, of the double solenoid is equal to the difference between the moments of the outer solenoid, M_2, and the inner solenoid, M_1,

$$M=M_1-M_2=0.1\pi IL[N_1R_1^2-N_2R_2^2]. \tag{4.4}$$

Thus for a net zero external moment M_1 shall equal M_2, i.e., the relation

$$N_1/N_2=(R_2/R_1)^2 \tag{4.5}$$

should hold. (If the solenoids under discussion consist of more than one layer of wire it can be shown that the radius, r, of a single-layered solenoid which has the same total number of turns and the same moment as a multi-layered coil is

$$r=[(r_o^2+r_or_i+r_i^2)/3]^{\frac{1}{2}}, \tag{4.6}$$

where r_o is the outer radius and r_i the inner radius of the multi-layered coils.)

5. Several double-solenoid neutralizing coils have been made and tested. Shown in Table I are the dimensions of a typical one of these.

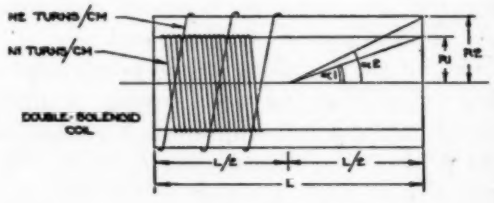

FIG. 2.

The constants of this double-solenoid neutralizing coil substituted into relations (4.3) and (4.4) gave:

$$H/I = 6.5 \text{ oersted/ampere},$$
$$M/I = 0.02 \text{ e.m.u./ampere}.$$

The above data can be compared with the data calculated for an ordinary solenoid of length, $L = 2.54$ cm, $2R = 1.59$ cm, and total number of turns $n = 16$. This gives nearly the same ratio of field intensity to current as the double solenoid, but it has an external moment about 150 times larger, i.e., for a single solenoid:

$$H/I = 6.73 \text{ oersted/ampere},$$
$$M/I = 3.17 \text{ e.m.u./ampere}.$$

6. Tests were made to determine the amount of external field associated with the double-solenoid neutralizing coil. With sufficient current in the double solenoid to produce a net internal field of 3.5 oersteds, the maximum external field along the axis was 0.006 oersteds measured one-half inch from the end of the coil. Use of a three-inch diameter toroidal neutralizing coil like that shown in Fig. 1 to produce a similar internal neutralizing field resulted in an external field, parallel to the internal field, of 0.02 oersteds measured at a point one-half inch to the side of the toroid.

MEASUREMENT OF MAGNETIC MOMENT

7. If a small magnet is placed in the center of a long solenoid, the flux linkage with the coil is proportional to the component of the magnetic moment of the magnet along the axis of the coil, M, in accordance with the relationship

$$M = \lambda/10K \text{ (e.m.u.)}, \qquad (7.1)$$

where λ is the flux linkage in Maxwell-turns of the magnet with the solenoid, and K is the coil constant of the solenoid in oersteds per ampere.*

TABLE I.

	Inner solenoid (1)	Outer solenoid (2)
Length (L)	2.54 cm	2.54 cm
Diameter ($2R$)	1.59 cm	2.46 cm
Total number of turns (n)	24	10

This is still true if the long solenoid is replaced by any coil which, when energized with current, produces a uniform field over the region occupied by the magnet. An ideal type of coil for the measurement of magnetic moments of small magnets is one which produces a uniform field inside the coil and zero field outside. This makes it possible for the magnet under test to be removed a relatively short distance from the coil to a position of practically zero flux linkage; if this is done suddenly it is possible to measure the change in flux by connecting a fluxmeter to the coil. The double solenoid very nearly approaches this type of ideal coil.

8. The magnetic moments of small magnets were measured by this method using a double-solenoid coil 20 inches long supported by a tube made of non-magnetic and non-conducting materials which held the test magnet in place. The tube was long enough so that the double solenoid could be easily slipped from over the magnet to a point 12 inches from the magnet where the linkage of flux was negligible. The change in flux linkage was measured with a fluxmeter, and values of magnetic moments were calculated from fluxmeter readings and the current constant of the double solenoid.

ACKNOWLEDGMENT

9. The authors are indebted to Mr. J. R. Wright of the Naval Ordnance Laboratory for the discussion of the measurement of magnetic moments of small magnets and to Dr. C. E. Barthel, Jr., of the Naval Ordnance Laboratory for his advice and assistance in this work.

* The relationship, $M = \lambda/10K$, is apparently not well known. It may be proved from the reciprocity relationship for mutual inductance. The magnet may be replaced by a small (secondary) coil of equal moment. The field, H_p, at the position of the magnet or small coil due to a current, I_p, in the primary coil may be expressed in the form

$$H_p = K_p I_p \text{ (e.m.u.)}, \qquad (7a)$$

which equation defines the coil constant K_p. Suppose the secondary coil has n_s turns and an area projected on a plane normal to the direction of the field of a_s square centimeters. Then λ_s, the flux linkage in the secondary due to the field, H_p, is

$$\lambda_s = H_p a_s n_s, \qquad (7b)$$

and the mutual inductance is

$$M_{sp} = \lambda_s/I_p = K_p a_s n_s. \qquad (7c)$$

The mutual inductance is also equal to flux linking the primary for unit current in the secondary

$$M_{sp} = \lambda_p/I_s = \lambda_p a_s n_s/M_s. \qquad (7d)$$

In the last equality, the current has been expressed in terms of the component of the moment of the coil in the direction of the field:

$$M_s = I_s a_s n_s. \qquad (7e)$$

Equation (7.1) may be obtained by combining (7c) and (7d) and expressing the coil constant K in oersteds per ampere instead of in e.m.u. ($K = 0.1K_p$). The moment of the magnet, of course, replaces the moment of the secondary coil.

The Class of Electron Lenses Which Satisfy Newton's Image Relation

R. G. E. HUTTER*
Division of Electron Optics Stanford University, California

(Received May 1, 1945)

Newton's image relations are satisfied for all electron lenses if object and image are located outside the region of the electromagnetic fields. It will be shown that a special class of electron lenses exists for which these relations hold for any object and image position. The results will be compared with those obtained by W. Glaser.

1. THE FOUR CARDINAL POINTS; IMAGE CONSTRUCTION

IT is known in light optics that a simple construction can be used to find the image of any object if a single set of the so-called "four cardinal points" of a lens is known. This construction is based on Newton's image relation and the magnification formula. The fact *that one and only one* set of cardinal points is necessary to construct the image of *any* object is the justification for the introduction of the concept of the cardinal points.

The four cardinal points are the two focal points z_{F_0}, z_{F_i}, and the two principal points z_{H_0}, z_{H_i}. The distances between the focal points and their corresponding principal points are the focal lengths f_0, f_i. The planes through the cardinal points perpendicular to the optical axis are called cardinal planes. There are two principal planes H_0, H_i,** and two focal planes F_0, F_i. z_0 and z_i are axial object and image coordinates respectively. $r(z_0)$ and $r(z_i)$ are points on the object and image respectively. r, z are considered to be cartesian coordinates. Focal lengths are positive quantities, and distances above the optical axis are positive, below the axis negative (see Fig. 1). Newton's image relation can then be written in the form:

$$(z_i - z_{F_i})(z_0 - z_{F_0}) = -f_0 f_i. \tag{1}$$

The magnification formula is:

$$M = r(z_i)/r(z_0) = f_0/(z_0 - z_{F_0}) = (-z_i - z_{F_i})/f_i. \tag{2}$$

Based on these two formulae is the following well-known image construction (see Fig. 1). A ray (the first principal ray) entering the lens from the object side parallel to the optical axis intersects H_i at A. The first locus of the image point I of the object point O is the straight line $\langle A z_{F_i} \rangle$. The second ray (the second principal ray) from O through the focal point z_{F_0} intersects H_0 at B. The second locus for the point I is the straight line through B parallel to the optical axis. I is the point of intersection of $\langle A z_{F_i} \rangle$ and the parallel line through B.

This simple construction gives correct results in all cases where object and image are outside the region which the actual lens occupies along the optical axis. Within the limits of the first order imagery the behavior of the lens is completely determined by a single set of cardinal points.

2. THE LIMITED USEFULNESS OF THE FOUR CARDINAL POINTS FOR THE IMAGE CONSTRUCTION

The significance of the four cardinal points is lost if these quantities become functions of the object position. We want to illustrate this for the case of a light optical system made up of infinitely many glass lenses arranged along an axis (see Fig. 2).

For any image only a finite number of lenses of the total system takes part in its formation. These lenses can be replaced by a single lens which has

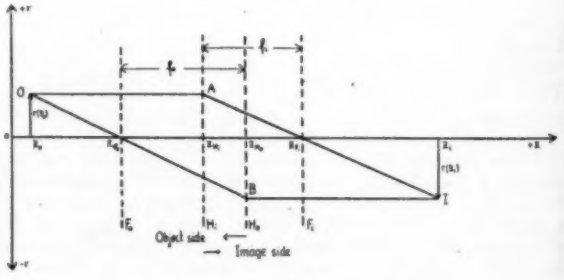

FIG. 1. Image construction.

* This paper is part of a dissertation submitted to the department of physics and the committee on graduate study of Stanford University in partial fulfillment of the requirements for the degree of doctor of philosophy. The author is now a research physicist with Sylvania Electric Products, Inc., New York. The paper was completed while he was research associate in the Division of Electron Optics at Stanford University.

** It is to be noted that in Fig. 1 the principal planes are crossed, that is, the object and image spaces overlap. The answer to the question whether *all* electron lenses have this property is not known to us. We will show later, however, that all lenses discussed in the present work must have crossed principal planes.

the same focal points and focal lengths as the combination lens. However, the number of lenses changes every time the object or the image changes position by one lens in the whole system so that every time this occurs a new set of cardinal points must be determined.

Any electron optical system whose axial field distribution is of infinite extent is analogous to the described glass lens arrangement. Object and image path which are the straight line portions of the total path cannot be distinguished anymore. It is hence impossible to speak of a one to one correspondence between the two sets of rays. The path of an electron is continuously curved.

It is therefore a surprising fact that nevertheless electron lenses exist which can be described in terms of a single set of cardinal points.

3. THE CONDITION FOR AN ELECTRON LENS TO HAVE A SINGLE SET OF CARDINAL POINTS

In order to determine the conditions for an electron lens to have a single set of cardinal points we start out with the paraxial ray differential equation. The solutions of this equation namely have now to satisfy Newton's image relation and the magnification formula, which in turn will impose conditions on the field quantities of the lens.

The paraxial ray differential equation of a combined electrostatic and magnetic electron lens is given by:

$$\phi(z)\cdot r''(z)+\tfrac{1}{2}\phi'(z)r'(z)+[\tfrac{1}{4}\phi''(z)+(e/8m)H^2(z)]r(z)=0, \quad (3)$$

where $\phi(z)$ is the electrostatic potential distribution along the optical axis, $H(z)$ is the magnetic field strength along the same axis, $r(z)$ is the function describing the distance of the electron from the optical axis* and e and m are the charge and mass of the electron respectively. The primed quantities are derivatives with respect to z.[1-4]

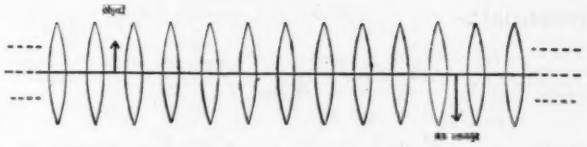

FIG. 2. Light optical system for which focal points and focal lengths are functions of object and image position.

To determine all the electron lenses whose optical properties are fully described in terms of a single set of cardinal points means mathematically to find all the functions $\phi(z)$ and $H(z)$ for which the solutions $r=r(z)$ of Eq. (3) satisfy Eqs. (1) and (2), where z_0 and z_i are any two successive roots of $r(z)$.

Equation (1) requires that a projective transformation shall hold between the successive roots of the path equation $r(z)$. This restriction on the solutions $r(z)$ of the differential Eq. (3) will lead us to a limited class of field functions $\phi(z)$ and $H(z)$. Equation (2) imposes a further restriction on the solutions of Eq. (3). This condition will put a restriction on the free choice of the focal lengths in case of purely magnetic or purely electrostatic lenses and will exclude the existence of combined lenses.

A solution of the stated problem becomes simpler if we apply a standard transformation to Eqs. (1)–(3). Let:

$$r(z)=y(z)\cdot\exp\left[-\frac{1}{4}\int\frac{\phi'(u)}{\phi(u)}du\right]=y(z)\cdot\phi^{-\frac{1}{4}}(z). \quad (4)$$

Then Eq. (3) becomes:[5]

$$y''(z)+\left\{\frac{e}{8m}\frac{H^2(z)}{\phi(z)}+\frac{3}{16}\left(\frac{\phi'(z)}{\phi(z)}\right)^2\right\}y(z)=0, \quad (5)$$

or:

$$y''(z)+\lambda(z)\cdot y(z)=0; \quad (6)$$

where:

$$\lambda(z)=\frac{e}{8m}\frac{H^2(z)}{\phi(z)}+\frac{3}{16}\left(\frac{\phi'(z)}{\phi(z)}\right)^2. \quad (7)$$

In case of a purely magnetic lens $\phi'(z)\equiv 0$, $\phi(z)=V=\text{const.}$, where V is the accelerating

* In case of the presence of a magnetic field the electron will describe a spiral motion about the optical axis. r, z, and a third coordinate—the angular displacement—would be necessary to describe completely the position of the electron everywhere in space. It is not essential in our problem, however, to consider the angular displacement of the electron. We can think of the r and z coordinates as a system of polar coordinates rotating with the electron around the z axis.

[1] E. Brueche and O. Scherzer, *Geometrische Elektronenoptik* (Verlagsbuchhandlung, Julius Springer, Berlin, 1934).

[2] L. M. Myers, *Electron Optics* (Chapman and Hall, Ltd., London, 1939).

[3] I. G. Maloff and D. W. Epstein, *Electron Optics in Television* (McGraw-Hill Book Company, Inc., New York, 1938), first edition.

[4] V. K. Zworykin and G. A. Morton, *Television* (John Wiley and Sons, Inc., New York, 1940).

[5] A. Recknagel, *AEG Forschungs-Institut, Jahrbuch. Sonderheft: Uebermikroskop* (Verlagsbuchhandlung Julius Springer, Berlin, 1940).

potential:

$$\lambda(z)=\lambda_m(z)=\frac{e}{8mV}H^2(z), \tag{8}$$

and in case of a purely electrostatic lens $H(z)\equiv 0$:

$$\lambda(z)=\lambda_e(z)=\frac{3}{16}\left(\frac{\phi'(z)}{\phi(z)}\right)^2. \tag{9}$$

We see that if the roots of the solution $r(z)$ satisfy Eq. (1) the roots of $y(z)$ will do likewise since both functions have the same roots ($r(z)$ differs from $y(z)$ only by an exponential function).

The condition (2) however becomes:

$$\frac{y(z_i)}{y(z_0)}=\frac{f_0}{z_0-z_{F_0}}\exp\left[\frac{1}{4}\int_{z_0}^{z_i}\frac{\phi'(u)}{\phi(u)}du\right]$$

$$=\frac{f_0}{z_0-z_{F_0}}\left(\frac{\phi(z_i)}{\phi(z_0)}\right)^{\frac{1}{4}}. \tag{10}$$

Now the problem is to find all the functions $\lambda=\lambda(z)$ defined by Eq. (7) for which the solutions $y=y(z)$ of the differential Eq. (6) satisfy the conditions (1) and (10).

4. DERIVATION OF A FUNCTIONAL RELATION FOR THE FIELD FUNCTION

The first step towards a solution of the problem will lead to a functional relation which $\lambda(z)$ must satisfy.

The general solution of the differential Eq. (6) can be written as:

$$y(z)=c_1\cdot y_1(z)+c_2\cdot y_2(z), \tag{11}$$

where $y_1(z)$ and $y_2(z)$ are two linearly independent solutions and c_1 and c_2 are two arbitrary constants. If $y(z)$ has a root at $z=z_0$ we can express $y(z)$ in the form:

$$y(z)=C\{y_2(z_0)\cdot y_1(z)-y_1(z_0)\cdot y_2(z)\}. \tag{12}$$

Let $z=z_i$ be the root following z_0; then:

$$y(z_i)=C\left\{y_2(z_0)\cdot y_1\left(z_{F_i}-\frac{f_0f_i}{z_0-z_{F_0}}\right)\right.$$

$$\left.-y_1(z_0)\cdot y_2\left(z_{F_i}-\frac{f_0f_i}{z_0-z_{F_0}}\right)\right\}=0, \tag{13}$$

assuming that Eq. (1) holds for z_0 and z_i. Hence:

$$\frac{y_1(z_0)}{y_2(z_0)}=\frac{y_1\left(z_{F_i}-\frac{f_0f_i}{z_0-z_{F_0}}\right)}{y_2\left(z_{F_i}-\frac{f_0f_i}{z_0-z_{F_0}}\right)}. \tag{14}$$

If we assume (1) to be true for all values $z=z_0$ we have:

$$\frac{y_1(z)}{y_2(z)}=\frac{y_1\left(z_{F_i}-\frac{f_0f_i}{z-z_{F_0}}\right)}{y_2\left(z_{F_i}-\frac{f_0f_i}{z-z_{F_0}}\right)}. \tag{15}$$

Differentiating both sides with respect to z, we get

$$\frac{1}{y_2^2(z)}=\frac{1}{y_2^2\left(z_{F_i}-\frac{f_0f_i}{z-z_{F_0}}\right)}\cdot\frac{f_0f_i}{(z-z_{F_0})^2}, \tag{16}$$

since:

$$y_2(z)\cdot y_1'(z)-y_1(z)\cdot y_2'(z)=\text{const.} \tag{17}$$

for all values of z (Wronskian determinant).

We obtain a similar expression for $y_1(z)$. Therefore $y(z)$ must satisfy the functional relation:

$$y(z)=\frac{z-z_{F_0}}{(f_0f_i)^{\frac{1}{2}}}\cdot y\left(z_{F_i}-\frac{f_0f_i}{z-z_{F_0}}\right) \tag{18}$$

$y(z)$ must also satisfy the differential Eq. (6), i.e.,

$$y''\left(z_{F_i}-\frac{f_0f_i}{z-z_{F_0}}\right)+\frac{(z-z_{F_0})^4}{(f_0f_i)^2}\cdot\lambda(z)$$

$$\cdot y\left(z_{F_i}-\frac{f_0f_i}{z-z_{F_0}}\right)=0. \tag{19}$$

We can therefore conclude that $\lambda(z)$ must satisfy the following functional relation:

$$\lambda(z)\cdot\frac{(z-z_{F_0})^4}{(f_0f_i)^2}=\lambda\left(z_{F_i}-\frac{f_0f_i}{z-z_{F_0}}\right). \tag{20}$$

5. DERIVATION OF AN ANALYTIC EXPRESSION FOR THE FIELD FUNCTION

In order to simplify the discussion of Eq. (20) we will first apply a number of transformations.

If we let:

$$z-z_{F_0}=x \quad\text{and}\quad z_{F_i}-z_{F_0}=d, \tag{21}$$

then Eq. (20) becomes:

$$L\left(d-\frac{f_0f_i}{x}\right)=\frac{x^4}{(f_0f_i)^2}\cdot L(x), \tag{22}$$

since

$$\lambda(z)=\lambda(x+z_{F_0})=L(x) \tag{23}$$

and

$$\lambda\left(z_{F_i}-\frac{f_0f_i}{z-z_{F_0}}\right)=\lambda\left(z_{F_0}+z_{F_i}-z_{F_0}-\frac{f_0f_i}{x}\right)$$

$$=L\left(z_{F_i}-z_{F_0}-\frac{f_0f_i}{x}\right)=L\left(d-\frac{f_0f_i}{x}\right). \tag{24}$$

Furthermore let

$$x=(f_0f_i)^{\frac{1}{2}}\cdot t \quad \text{and} \quad \frac{d}{(f_0f_i)^{\frac{1}{2}}}=E; \tag{25}$$

then Eq. (22) becomes:

$$M\left(E-\frac{1}{t}\right)=t^4\cdot M(t), \tag{26}$$

since

$$L(x)=L[(f_0f_i)^{\frac{1}{2}}\cdot t]=M(t), \tag{27}$$

and

$$L\left(d-\frac{f_0f_i}{x}\right)=L\left[d-\frac{f_0f_i}{(f_0f_i)^{\frac{1}{2}}t}\right]$$

$$=L\left[(f_0f_i)^{\frac{1}{2}}\left(\frac{d}{(f_0f_i)^{\frac{1}{2}}}-\frac{1}{t}\right)\right]=M\left(E-\frac{1}{t}\right). \tag{28}$$

We see that a function satisfying the functional relation (26) is given by:

$$N(t)=\frac{1}{(1-Et+t^2)^2}, \tag{29}$$

which can easily be verified. This $N(t)$ is essentially the function representing the magnetic field distribution along the optical axis whose lens properties were discussed by W. Glaser.[6]

If $M(t)$ is assumed to be another solution of Eq. (26) the ratio $M(t)/N(t)$ defines a new function $f(t)$:

$$f(t)=M(t)/N(t) \tag{30}$$

and $f(t)$ must satisfy the functional relation

$$f\left(E-\frac{1}{t}\right)=f(t); \tag{31}$$

if $f(t)$ are all the solutions of Eq. (31) the solutions of Eq. (26) are then given by

$$M(t)=N(t)\cdot f(t)=\frac{1}{(1-Et+t^2)^2}f(t). \tag{32}$$

The fixed points of the transformation

$$t'=E-1/t \tag{33}$$

are

$$\alpha=\frac{E}{2}+\left(\frac{E^2}{4}-1\right)^{\frac{1}{2}}, \quad \beta=\frac{E}{2}-\left(\frac{E^2}{4}-1\right)^{\frac{1}{2}}. \tag{34}$$

Depending on the value of E there are either two real, one real or two conjugate complex fixed points.
If:

1. $E>2$, α and β are real quantities;
2. $E=2$, $\alpha=\beta$ and real; (35)
3. $0<E<2$, α and β are conjugate complex quantities.

Let

$$t=(\beta u-\alpha)/(u-1)$$

and

$$t'=(\beta u'-\alpha)/(u'-1), \tag{36}$$

or

$$u=(t-\alpha)/(t-\beta) \quad \text{and} \quad u'=(t'-\alpha)/(t'-\beta); \tag{37}$$

then

$$u'=(t'-\alpha)/(t'-\beta)=\rho(t-\alpha)/(t-\beta)=\rho\cdot u, \tag{38}$$

where

$$\rho=\beta/\alpha \tag{39}$$

since $t'=\infty$ for $t=0$, $(\alpha/\beta)\cdot\rho=1$.

Then $f(t)$ becomes a function $F(u)$ which satisfies the functional relation:

$$F(\rho\cdot u)=F(u). \tag{40}$$

If $E>2$, ρ is a real quantity and $0<\rho<1$. For this condition the only solution of Eq. (40) is:

$$F(u)=\text{const.}$$

hence

$$f(t)=\text{const.}$$

and the only solution of Eq. (26) is given by Eq. (29) multiplied by a constant

$$M(t)=\text{const.}/(1-Et+t^2)^2. \tag{41}$$

This function, however, does not lead to fields which can be realized in practice.

The same is true for the case $E=2$; we therefore can omit a detailed discussion of these two cases.

[6] W. Glaser, Zeits. f. Physik 117, 285 (1941).

If $0<E<2$, we can write E as:

$$E=2\cdot\cos\varphi \tag{42}$$

where

$$0<\varphi<\pi/2.$$

Then

$$\alpha=\cos\varphi+i\sin\varphi=e^{i\varphi},$$
$$\beta=\cos\varphi-i\sin\varphi=e^{-i\varphi}, \tag{43}$$

and the value of ρ in Eq. (38) is given by

$$\rho=\beta/\alpha=e^{-2i\varphi},\quad |\rho|=1 \tag{44}$$

u and u' are now complex quantities. The t axis is transformed into the unit circle and $F(u)$ is a real and continuous function on the unit circle $|u|=1$. Since

$$F(\rho\cdot u)=F(u). \tag{45}$$

We expand $F(u)$ in a Fourier series

$$F(u)=F(e^{i\theta})=\sum_{n=-\infty}^{n=+\infty}a_n e^{in\theta}. \tag{46}$$

Then Eq. (45) is:

$$\sum_{n=-\infty}^{n=+\infty}a_n e^{in(\theta-2\varphi)}=\sum_{n=-\infty}^{n=+\infty}a_n e^{in\theta}. \tag{47}$$

Hence

$$a_n(e^{-2in\varphi}-1)=0. \tag{48}$$

(i) If φ/π is irrational, a_n must be equal to zero for all n except $n=0$ and $f(t)=\text{const}$. Therefore $N(t)$ given by Eq. (29) is the only solution.

(ii) If φ/π is rational, φ/π can be written as p/q where $p>0$, $q>0$ and p and q are relative prime integers. Then:

$$e^{-2in\pi p/q}=1, \tag{49}$$

for all $n=mq$, m being an integer. Since

$$u=(t-\alpha)/(t-\beta)=e^{i\theta},$$
$$e^{in\theta}=s^{imq\theta}=((t-\alpha)/(t-\beta))^{mq};$$

and

$$F(u)=f(t)=\sum_{m=-\infty}^{m=+\infty}c_m((t-\alpha)/(t-\beta))^{mq}, \tag{50}$$

with

$$\alpha=\beta=e^{i\varphi}\quad\text{and}\quad c_{-m}=\bar{c}_m,$$

where the c_m are any arbitrary constants.

We therefore obtain the following complete solution of the functional relation (26)

$$M(t)=\frac{1}{(1-Et+t^2)^2}\times\sum_{m=0}^{\infty}c_m\left\{\left(\frac{t-\alpha}{t-\beta}\right)^{mq}+\left(\frac{t-\beta}{t-\alpha}\right)^{mq}\right\}, \tag{51}$$

which is a real function (two conjugate complex terms of the series (51) are combined to give a real expression).

If we reverse the substitutions (21) and (25) the expression for $\lambda(z)$—the function which satisfies Eq. (20)—becomes:

$$\lambda(z)=\sum_{m=0}^{\infty}\frac{c_m}{(f_0f_i)^2}\frac{\{z-\frac{1}{2}(z_{Fi}+z_{F0})-[\frac{1}{4}(z_{Fi}-z_{F0})^2-f_0f_i]^{\frac{1}{2}}\}^{2mq}+\{z-\frac{1}{2}(z_{Fi}+z_{F0})+[\frac{1}{4}(z_{Fi}-z_{F0})^2-f_0f_i]^{\frac{1}{2}}\}^{2mq}}{[(z-z_{F0})^2-(z_{Fi}-z_{F0})(z-z_{F0})+f_0f_i]^{mq+2}}. \tag{52}$$

We should like to remark that our distinction between rational and irrational values of φ/π loses its meaning in physical applications and is mentioned only to give the complete mathematical solution of the problem.

6. THE RESTRICTING INFLUENCE OF THE MAGNIFICATION FORMULA

So far we have used only the lens equation as a restricting condition imposed on the solutions of the paraxial ray equation. In terms of the solution of Eq. (6) we had written the magnification formula as:

$$y(z_i)/y(z_0)=[f_0/(z_0-z_{F0})](\phi(z_i)/\phi(z_0))^{\frac{1}{2}} \tag{10}$$

or

$$y(z)=\frac{z-z_{F0}}{f_0}\left(\frac{\phi(z_{Fi}-f_0f_i/(z-z_{F0}))}{\phi(z)}\right)^{-\frac{1}{2}}\cdot y\left(z_{Fi}-\frac{f_0f_i}{z-z_{F0}}\right). \tag{53}$$

If we compare this equation with Eq. (18) which we derived by using the lens equation only

$$y(z)=\frac{z-z_{F0}}{(f_0f_i)^{\frac{1}{2}}}\cdot y\left(z_{Fi}-\frac{f_0f_i}{z-z_{F0}}\right) \tag{18}$$

we see that

$$(f_0f_1)^{\frac{1}{2}}=f_0\left(\frac{\phi(z_{Fi}-f_0f_i/(z-z_{F0}))}{\phi(z)}\right)^{\frac{1}{2}}, \tag{54}$$

or

$$\frac{f_i}{f_0}=\left[\frac{\phi(z_{F_i}-f_0f_i/(z-z_{F_0}))}{\phi(z)}\right]^{\frac{1}{2}}. \tag{55}$$

This expression must be independent of z, otherwise the pair of focal lengths would become a function of the object position.

We will distinguish three cases: (1) Purely magnetic lens. $\phi(z)=$const in this case. Therefore $f_0=f_i$. (2) Purely electrostatic lens. $H(z)\equiv 0$ in this case. It is:

$$\lambda(z)=\lambda_e(z)=\tfrac{3}{16}(\phi'(z)/\phi(z))^2,$$

and $\phi'(z)/\phi(z)$ must satisfy the following functional relation:

$$\frac{\phi'(z_{F_i}-f_0f_i/(z-z_{F_0}))}{\phi(z_{F_i}-f_0f_i/(z-z_{F_0}))}=\frac{(z-z_{F_0})^2}{f_0f_i}\frac{\phi'(z)}{\phi(z)},$$

or

$$\frac{\phi'(z_{F_i}-f_0f_i/(z-z_{F_0}))}{\phi(z_{F_i}-f_0f_i/(z-z_{F_0}))}\cdot\frac{f_0f_i}{(z-z_{F_0})^2}=\frac{\phi'(z)}{\phi(z)}. \tag{56}$$

Integration with respect to z gives:

$$\phi\left(z_{F_i}-\frac{f_0f_i}{z-z_{F_0}}\right)=\text{const.}\ \phi(z). \tag{57}$$

Hence f_i/f_0 is independent of z. (3) Combined electrostatic and magnetic lens.

The requirement that f_i/f_0 must be constant leads to Eq. (56) by the differentiation of Eq. (55). This means that the term $\frac{3}{16}(\phi'(z)/\phi(z))^2$ in Eq. (7) must be of the form given by Eq. (52). But it is then impossible to satisfy the condition that the total expression for $\lambda(z)$ is of the same form since $\phi(z)$ appears in the denominator of the term $(e/8m)(H^2(z)/\phi(z))$. A combined lens satisfying the lens equation and the magnification formula cannot therefore exist.

7. PHYSICAL SIGNIFICANCE OF THE CONDITION FOR THE PARAMETER E

Solutions of the problem exist only for the case that $0<E<2$ where E was defined by

$$d/(f_0f_i)^{\frac{1}{2}}=E<2. \tag{25}$$

For this condition:

$$d^2/4<f_0f_i. \tag{58}$$

It can be seen from Fig. 1 that $f_0+f_i<d$ is the condition that the principal planes are not crossed. This inequality contradicts, however, Eq. (58) as we can readily prove by the following consideration:

It is $f_0+f_i<d$, hence

$$\frac{f_0^2+2f_0f_i+f_i^2}{4}<\frac{d^2}{4};$$

or

$$\frac{d^2}{4}>\frac{f_0f_i}{2}+\frac{f_0^2}{4}+\frac{f_i^2}{4}+\frac{f_0f_i}{2}-\frac{f_0f_i}{2},$$

$$\frac{d^2}{4}>f_0f_i+\frac{(f_0-f_i)^2}{4},$$

since

$$\frac{(f_0-f_i)^2}{4}\geqq 0,\quad \frac{d^2}{4}\geqq f_0f_i,$$

but Eq. (58) required $d^2/4<f_0f_i$. Therefore all the lenses given by the function $\lambda(z)$ have crossed principal planes.

Within the range $0<E<2$ it is possible to choose E arbitrarily. Since:

$$\frac{d}{(f_0f_i)^{\frac{1}{2}}}=\frac{z_{F_i}-z_{F_0}}{(f_0f_i)^{\frac{1}{2}}}=E,$$

d, z_{F_0}, and $f_0\cdot f_i$ can be chosen so that $0<E<2$. The ratio f_0/f_i and thus the individual focal lengths are then determined by Eq. (55). The field function of the lens having these lens parameters z_{F_0}, z_{F_i}, f_0, f_i is then determined by Eq. (52).

8. SIMPLE EXAMPLES OF ELECTRON LENSES OF THE TYPE DISCUSSED

It remains to be shown that certain of the class of field functions given by Eq. (52) are physically realizable. We will list a number of simple field functions constituting proper lenses which belong to this class.

A. Magnetic Lenses

(1) $H(z)=1/(1+z^2)$, (3) $H(z)=z/(1+z^2)^{\frac{1}{2}}$,

(2) $H(z)=1/(1+z^2)^{\frac{1}{2}}$, (4) $H(z)=z/(1+z^2)^2$,

(5) $H(z)=(z^4+1)^{\frac{1}{2}}/(1+z^2)^2$.

Sketches of some of these field functions are shown in Figs. 3 to 5.

B. Electrostatic Lenses

(1) $\phi(z)=\exp\left[c\arctan z\right]$,

(2) $\phi(z)=\exp\left[c\dfrac{1}{(1/z^2+1)^{\frac{1}{2}}}\right]$,

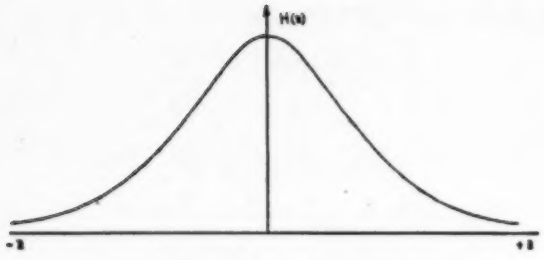

FIG. 3. Sketch of $H(z)=1/(1+z^2)$.

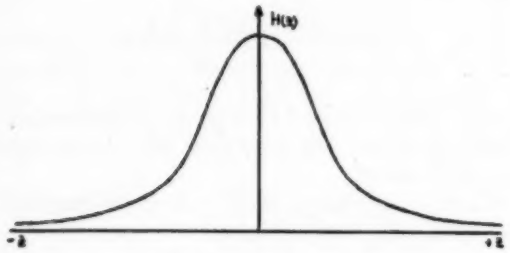

FIG. 4. Sketch of $H(z)=1/(1+z^2)^{\frac{1}{2}}$.

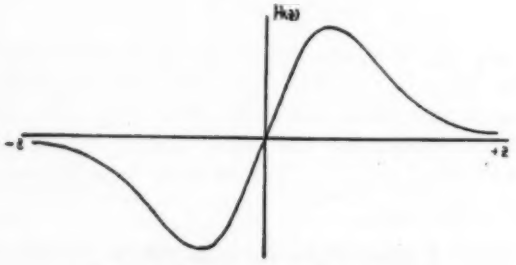

FIG. 5. Sketch of $H(z)=z/(1+z^2)^{\frac{1}{2}}$.

$$(3)\quad \phi(z)=k\cdot\exp\left[c\frac{1}{(z^2+1)^{\frac{1}{2}}}\right],$$

$$(4)\quad \phi(z)=k\cdot\exp\left(c\frac{1}{1+z^2}\right).$$

Some of these axial potential distributions are shown in Figs. 6 and 7.

9. COMPARISON OF THE RESULTS OBTAINED HERE WITH THOSE FOUND BY W. GLASER[7]

The investigations presented here were concluded when a paper published by W. Glaser in 1941 came to the attention of the author. In this paper Glaser obtained a solution of the same problem by a different mathematical approach. The form of his solution also has a different appearance from the one given here so that it is worthwhile comparing the two results. Glaser's solution is as follows:

An image equation of the type of geometrical light optics:

$$z_1=(az+b)/(cz+d) \tag{59}$$

exists between object and image coordinates z and z_1 if and only if the electromagnetic field along the axis has the general form

$$\frac{3}{16}\left(\frac{\phi'(z)}{\phi(z)}\right)^2+\frac{e}{8m}\frac{H^2(z)}{\phi(z)}=\frac{1}{4}\frac{(a-d)^2+4bc}{[cz^2+(d-a)z-b]^2}[1-J(\zeta)], \tag{60}$$

where $J(\zeta)$ is a periodic function of the argument

$$\zeta=\operatorname{arc\,cot}\frac{2cz-a+d}{[-[(a-d)^2+4bc]]^{\frac{1}{2}}}, \tag{61}$$

with the period

$$\gamma=\operatorname{arc\,cot}\frac{a+d}{[-[(a-d)^2+4bc]]^{\frac{1}{2}}}. \tag{62}$$

We can write Eq. (1) also in the form

$$z_i=\frac{z_{F_i}z-[f_0f_i+z_{F_i}z_{F_0}]}{z-z_{F_0}}. \tag{63}$$

Comparison with Eq. (59) yields:

$$a=z_{F_i},\quad b=-[f_0f_i+z_{F_i}z_{F_0}],\quad c=1,\quad d=-z_{F_0}, \tag{64}$$

and

$$E=(z_{F_i}-z_{F_0})/(f_0f_i)^{\frac{1}{2}}=(a+d)/(ad-bc)^{\frac{1}{2}}; \tag{65}$$

furthermore

$$\cot\zeta=\frac{2z-(z_{F_i}+2z_{F_0})+z_{F_0}}{[-(z_{F_i}-z_{F_0})^2-4f_0f_i]^{\frac{1}{2}}}=\frac{t-E/2}{[-(E^2/4-1)]^{\frac{1}{2}}}, \tag{66}$$

$$\cot\gamma=\frac{z_{F_i}-z_{F_0}}{[-(z_{F_i}-z_{F_0})^2-4f_0f_i]^{\frac{1}{2}}}=\frac{E/2}{[-(E^2/4-1)]^{\frac{1}{2}}}, \tag{67}$$

[7] W. Glaser and E. Lammel, Ann. d. Physik [5] 40, 367 (1941).

and

$$\lambda(t)=\frac{E^2/4-1}{f_0 f_i}\,\frac{1}{[t^2-Et+1]^2}\times\left[1-J\left(\operatorname{arc\,cot}\frac{t-E/2}{(1-E^2/4)^{\frac{1}{2}}}\right)\right], \quad (68)$$

where J has the period:

$$\gamma=\operatorname{arc\,cot}\frac{E/2}{(1-E^2/4)^{\frac{1}{2}}}. \quad (69)$$

It was:

$$t=(\beta u-\alpha)/(u-1), \quad (36)$$

with

$$\alpha=\frac{E}{2}+\left(\frac{E^2}{4}-1\right)^{\frac{1}{2}},\quad \beta=\frac{E}{2}-\left(\frac{E^2}{4}-1\right)^{\frac{1}{2}}; \quad (34)$$

hence

$$\frac{t-E/2}{(1-E^2/4)^{\frac{1}{2}}}=-i\frac{u+1}{u-1}; \quad (70)$$

therefore

$$J\left[\operatorname{arc\,cot}\frac{t-E/2}{(1-E^2/4)^{\frac{1}{2}}}\right]=J\left(\operatorname{arc\,cot}\left(-i\frac{u+1}{u-1}\right)\right), \quad (71)$$

since

$$\operatorname{arc\,cot}\frac{1}{z}=\frac{i}{2}\ln\frac{1-iz}{1+iz},$$

$$J\left(\operatorname{arc\,cot}\left(-i\frac{u+1}{u-1}\right)\right)=J\left(\frac{i}{2}\ln\frac{1+\dfrac{u+1}{u-1}}{1-\dfrac{u-1}{u+1}}\right)=J\left(\frac{i}{2}\ln u\right), \quad (72)$$

hence

$$\zeta=\frac{i}{2}\ln u. \quad (73)$$

From Eq. (42)

$$2\cos\varphi=E, \quad (42)$$

$$\cos^2\varphi=E^2/4,\quad \sin^2\varphi=1-E^2/4,$$

$$\cot\varphi=\cos\varphi/\sin\varphi=(E/2)/(1-E^2/4)^{\frac{1}{2}}$$

hence

$$\varphi=\gamma. \quad (74)$$

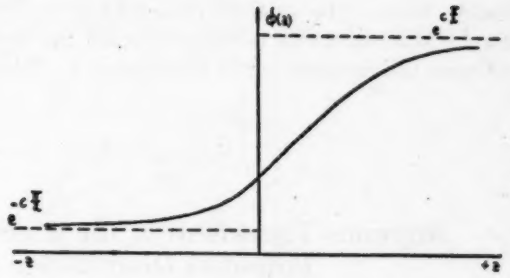

FIG. 6. Sketch of $\phi(z)=e^{c\cdot\arctan z}$.

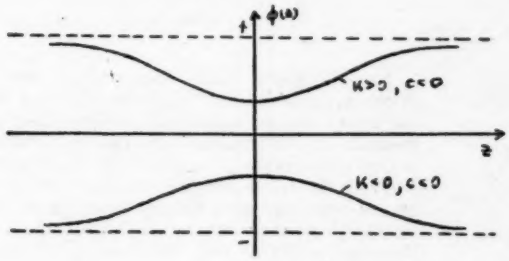

FIG. 7. Sketch of $\phi(z)=k\cdot\exp\left[c\frac{1}{1+z^2}\right]$.

Therefore with Eq. (44)

$$\rho=e^{-2i\varphi} \quad (44)$$

or

$$\gamma=(i/2)\ln\rho. \quad (75)$$

$J(\zeta)$ is, according to Glaser, a periodic function with the period γ; it can therefore be expanded in a Fourier series

$$\sum_{n=-\infty}^{+\infty}a_n e^{in(\zeta+\gamma)}=\sum_{n=-\infty}^{+\infty}a_n e^{in\zeta}; \quad (76)$$

this means

$$e^{in\gamma}=1=\exp\left[-2\pi i\left(\frac{\gamma}{\pi}\cdot\frac{n}{2}\right)\right], \quad (77)$$

hence

$$n=m\cdot 2q. \quad (78)$$

This leads to:

$$J(\zeta)=\sum a_{2mq}e^{mq\ln u}=\sum a_{2mq}u^{mq}=M(t). \quad (79)$$

Both results of the same problem are therefore identical despite their difference in formal appearance.

ACKNOWLEDGMENT

The present work was carried out while the author was connected with the Division of Electron Optics at Stanford University. For

valuable advice the author is indebted to Professor L. Marton under whose direction the work was done. Discussions with Professors G. Szego and G. Polya of Stanford University were helpful in obtaining solutions of a number of mathematical problems.

Rigorous Treatment of the Electrostatic Immersion Lens Whose Axial Potential Distribution Is Given by: $\phi(z)=\phi_0 e^{K \arctan z}$

R. G. E. HUTTER*
Division of Electron Optics, Stanford University, California
(Received May 9, 1945)

Investigations have shown that the electrostatic immersion lens, given by the axial potential distribution $\phi(z)=\phi_0 \exp(K \arctan z)$, is the simplest electrostatic lens for which Newton's image equations are satisfied throughout the field. The general solution of the paraxial ray differential equation is derived. Based on this solution and the fact that Newton's image equations are satisfied, exact expressions are derived for all important optical quantities such as the focal lengths, the location of the four cardinal points, magnification, the object-image relation, and the spherical and chromatic aberration for any object position. The optimum positions of an object are determined so that the chromatic and spherical aberrations are reduced to a minimum. All quantities are represented graphically. A plot of the equipotential lines in space is given for a special value of the voltage ratio.

ELECTRON lenses of rotational symmetry are uniquely described by their axial field distribution. The field everywhere in space, the electron paths, as well as all optical quantities, can be expressed in terms of the axial field function and its derivatives. Special conditions imposed on lenses, such as minimum of aberrations or the validity of Newton's image equations, can be interpreted as conditions for the axial field function. We therefore have frequently, at the start of the actual investigation of an electrostatic lens, the axial potential as the basic quantity.

In electrostatic electron optics it often occurs, however, that lenses given in terms of their axial potential function are hard to realize in practice by means of simple electrode configurations. Singular points of the potential cause this quantity to be multivalued in the normal cartesian plane, and it does not help the physicist that the function becomes single-valued in the Riemannian plane. No mechanic can build lenses in this kind of space. In such cases it is only possible to approximate to the desired potential distribution. The problem whether or not the axial potential function

$$\phi(z)=\phi_0 e^{K \arctan (z/a)} \tag{1}$$

will give rise to a multivalued potential distribution in space can be solved by a computation of the equipotential surfaces. By inspection of such an equipotential plot it is then also possible to determine an electrode configuration producing exactly or approximately the axial potential distribution of Eq. (1). This distribution is plotted in Fig. 1 for some values of the parameters ϕ_0, κ and a. It is compared with that for a special cylinder lens.

1. THE EQUIPOTENTIAL PLOT

A first glance at the function $\phi(z+ir)$ reveals that it has a logarithmic singularity at $r=1$, $z=0$, since

$$\arctan u=\frac{i}{2}\ln\frac{i+u}{i-u}=\frac{i}{2}\ln\frac{z+i(1+r)}{-z+i(1-r)}$$

if

$$u=z+ir.$$

This shows that the series

$$\varphi(z,r)=\sum_{v=0}^{\infty}\frac{(-1)^v}{(v!)^2}\phi^{(2v)}(z)\left(\frac{r}{2}\right)^{2v},$$

commonly used for the determination of the

* This paper is part of a dissertation submitted to the department of physics and the committee on graduate study of Stanford University in partial fulfillment of the requirements for the degree of doctor of philosophy. The author is now a research physicist with Sylvania Electric Products Inc., New York. The paper was completed while he was research associate in the Division of Electron Optics at Stanford University.

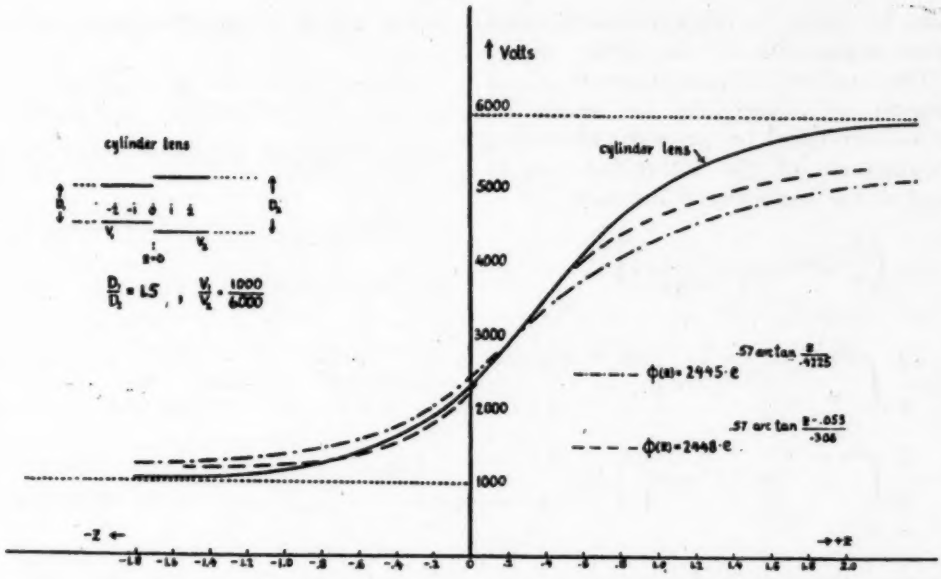

FIG. 1. Axial potential functions.

potential distribution in space, will converge only for values of r with $r < r_0$, where

$$r_0 = (1+z^2)^{\frac{1}{2}}.$$

For values of r near the limits of convergence a great many terms are necessary before it is permissible to break off the series for numerical computation. If $r > r_0$ the series is divergent, so that no information can be gained from it about the character of the potential function.

Another method makes use of the solution of Laplace's equation in the form of a definite integral:

$$\varphi(z,r) = \frac{1}{2\pi}\int_0^{2\pi} \phi(z+ir\cos\alpha)\,d\alpha = \frac{2}{\pi}\int_0^{\pi/2} Re\{\phi(z+ir\cos\alpha)\}\,d\alpha.$$

In our special case we have:

$$K \arctan(z+ir\cos\alpha) = \frac{Ki}{2}\ln\frac{i(1+r\cos\alpha)+z}{i(1-r\cos\alpha)+z},$$

so that:

$$e^{K\arctan(z+ir\cos\alpha)} = \exp\left\{\frac{Ki}{2}\left[\ln\left(\frac{z^2+(1+r\cos\alpha)^2}{z^2+(1-r\cos\alpha)^2}\right)^{\frac{1}{2}} - i\arctan\frac{2z}{(1-r^2\cos^2\alpha)-z^2}\right]\right\}.$$

The real part of this function is then:

$$Re[e^{K\arctan(z+ir\cos\alpha)}] = \left\{\exp\left[\frac{K}{2}\arctan\frac{2z}{(1-r^2\cos^2\alpha)z^2}\right]\right\} \times\cos\left[\frac{K}{4}\ln\frac{z^2+(1+r\cos\alpha)^2}{z^2+(1-r\cos\alpha)^2}\right]; \quad (2)$$

or

$$Re[e^{K\arctan(z+ir\cos\alpha)}] = \left\{\exp\left[\frac{K}{2}\left(\arctan\frac{z}{1-r\cos\alpha} + \arctan\frac{z}{1+r\cos\alpha}\right)\right]\right\} \times\cos\left[\frac{K}{4}\ln\frac{z^2+(1+r\cos\alpha)^2}{z^2+(1-r\cos\alpha)^2}\right],$$

since:

$$\arctan a+\arctan b = \arctan\frac{a+b}{1-ab},$$

$$\arctan\frac{z}{1-r\cos\alpha}+\arctan\frac{z}{1+r\cos\alpha} = \arctan\frac{2z}{(1-r^2\cos^2\alpha)-z^2}.$$

The integral form of the potential function permits the application of the process of analytic

continuation, by which the potential distribution in the region inaccessible by the series can be obtained. The analytic continuation of $\varphi(z, r)$ into the region not covered by the series can readily be accomplished by properly observing the multivaluedness of the arctan-function in the exponent of the exponential function:

$$\exp\left(\frac{K}{2}\arctan\frac{z}{1-r\cos\alpha}\right)$$

of Eq. (2). If z is positive and $r>(1+z^2)^{\frac{1}{2}}$ then:

$$\arctan\frac{z}{1-r\cos\alpha}=2\pi-\arctan\frac{z}{|1-r\cos\alpha|},$$

and for negative values of z and $r>(1+z^2)^{\frac{1}{2}}$:

$$\arctan\frac{z}{1-r\cos\alpha}=-2\pi+\arctan\left|\frac{z}{1-r\cos\alpha}\right|.$$

If $z=0$ it follows that for $r>1$:

$$\varphi(+0,r)=\frac{2}{\pi}\int_0^{\arccos 1/r} e^{K\pi/2}\cos\left[\frac{K}{2}\ln\frac{1+r\cos\alpha}{-1+r\cos\alpha}\right]d\alpha+\frac{2}{\pi}\int_{\arccos 1/r}^{\pi/2}\cos\left[\frac{K}{2}\ln\frac{1+r\cos\alpha}{1-r\cos\alpha}\right]d\alpha, \quad (3)$$

$$\varphi(-0,r)=\frac{2}{\pi}\int_0^{\arccos 1/r} e^{-K\pi/2}\cos\left[\frac{K}{2}\ln\frac{1+r\cos\alpha}{-1+r\cos\alpha}\right]d\alpha+\frac{2}{\pi}\int_{\arccos 1/r}^{\pi/2}\cos\left[\frac{K}{2}\ln\frac{1+r\cos\alpha}{1-r\cos\alpha}\right]d\alpha. \quad (4)$$

This shows that the potential function is univalued in the plane the physicist is interested in, except for the portion of the r axis for values from $r=1$ to $r=\infty$. It is through the cut along this line that the function $\varphi=\varphi(z,r)$ disappears into the other sheets of the Riemannian plane. In the cartesian plane $\varphi(z,r)$ is double valued along this line.

For the application of formulae (3) and (4), which determine the potential along the r axis, it is helpful to know the values of α for which

$$\cos\left[\frac{K}{2}\ln\frac{1+r\cos\alpha}{\pm1\mp r\cos\alpha}\right] \quad (5)$$

becomes indefinite, has maxima and minima and zeros. For any z these values are determined by:

$$\frac{K}{4}\ln\frac{(1+r\cos\alpha)^2+z^2}{(1-r\cos\alpha)^2+z^2}=n\cdot\frac{\pi}{2}.$$

$n=2m+1$, $m=0, 1, 2, \cdots$ zeros;

$n=2m$, $m=0, 1, 2, \cdots$ maxima and minima.

Equation (5) leads to:

$$\cos\alpha_{1,2}=\frac{1}{r}\left[\coth\frac{\pi n}{K}\pm\left(\coth^2\frac{\pi n}{K}-(z^2+1)\right)^{\frac{1}{2}}\right].$$

For the case $z=0$ the zeros are therefore given by:

$$\cos\alpha_{1,2}=\frac{1}{r}\frac{\cosh\frac{(2m+1)\pi}{K}\pm1}{\sinh\frac{(2m+1)\pi}{K}}.$$

The integrand becomes indetermined for:

$$\alpha_c=\arccos\frac{1}{r}, \quad \text{for } r\geqq1,\ 0\leqq\alpha_c<\frac{\pi}{2}.$$

The general behavior of the integrand is illustrated in Fig. 2. It is noteworthy that the zeros of the integrand appear to be clustered around

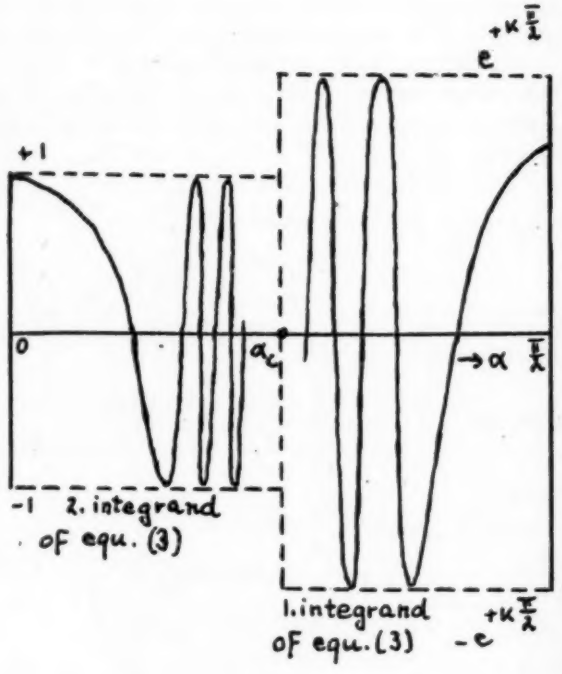

FIG. 2. The integrands in Eq. (3).

the indeterminate point (for $z=0$) as given by:

$$\cos \alpha_e = 1/r,$$

so that the areas under the oscillatory part of the curve could very well be neglected when performing the integration by means of the planimeter.

After we had obtained a general picture of the behavior of the equipotential lines in space another method was suggested to us by Professor W. W. Hansen of Stanford University. Since this method reduced the amount of computational work considerably, it was used to determine the actual potential plot of Fig. 3.

The solution of Laplace's equation in spherical coordinates with rotational symmetry about the z axis can be written as:

$$\varphi_1(z, \rho) = \sum_{n=0}^{\infty} a_n \rho^n P_n(\cos\theta) \qquad \text{for } \rho < 1,\ 0 \leqq \theta \leqq \pi;$$

$$\varphi_2(z, \rho) = \sum_{n=0}^{\infty} b_n \frac{1}{\rho^{n+1}} P_n(\cos\theta) + \phi_{+\infty} \qquad \text{for } \rho > 1,\ 0 \leqq \theta \leqq \pi/2;$$

$$\varphi_3(z, \rho) = \sum_{n=0}^{\infty} c_n \frac{1}{\rho^{n+1}} P_n(\cos\theta) + \phi_{-\infty} \qquad \text{for } \rho > 1,\ \pi/2 \leqq \theta \leqq \pi;$$

where ρ is the distance from the origin of the spherical system of coordinates, θ the angle between the positive z axis and ρ, and P_n is the nth Legendre polynomial. (Fig. 4.)

The coefficients a_n, b_n, c_n must be determined so that the potential functions φ_1, φ_2, φ_3 represent the axial potential

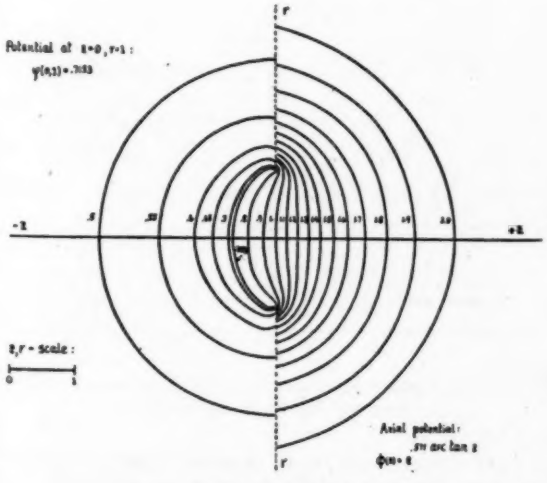

FIG. 3. Equipotential lines.

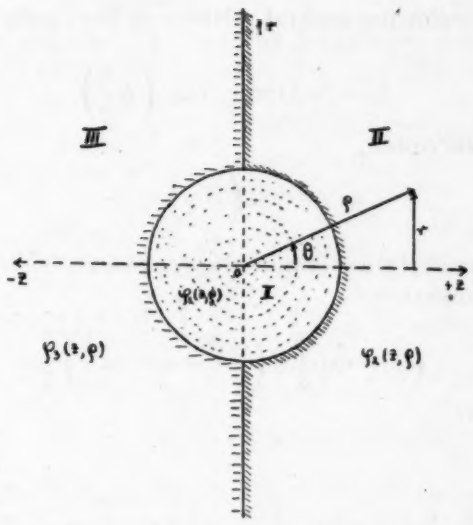

FIG. 4. Spherical coordinates (with rotational symmetry about the z axis).

$$\phi(z) = e^{K \arctan z} \tag{6}$$

in their respective regions.

The coefficients a_n will be obtained by an expansion of our axial potential in a power series of z. We get:

$$(n+1)!a_{n+1} = Ka_n n! - n(n-1)a_{n-1}(n-1)!$$

$$n = 0, 1, 2, \ldots; \quad a_0 = 1.$$

The first 10 coefficients $(n!a_n)$ are given by:

$$a_0 = 1, \quad a_1 = K, \quad 2a_2 = K^2 - 2, \quad 3!a_3 = K^3 - 8K,$$
$$4!a_4 = K^4 - 20K^2 + 24, \quad 5!a_5 = K^5 - 40K^3 + 184K,$$
$$6!a_6 = K^6 - 70K^4 + 784K^2 - 720,$$
$$7!a_7 = K^7 - 112K^5 + 2464K^3 - 8448K,$$
$$8!a_8 = K^8 - 168K^6 + 6384K^4 - 52352K^2 + 40320,$$
$$9!a_9 = K^9 - 240K^7 + 14448K^5 - 249760K^3 + 648576K.$$

In order to determine the coefficients b_n we expand

$$\phi(z) = \exp\left[K\left(\frac{\pi}{2} - \arctan\frac{1}{z}\right)\right], \quad \infty > z \geqq 1$$

in a series of powers of $1/z$. We get for the coefficients in:

$$\phi(z) = \phi_{+\infty} + b_0 \cdot \frac{1}{z} + b_1 \cdot \frac{1}{z^2} + b_2 \cdot \frac{1}{z^3} + \cdots + b_n \frac{1}{z^{n+1}} + \cdots$$

the following general relation to the coefficients a_n:

$$b_n=(-1)^n a_{n+1}\cdot\exp\left(K\frac{\pi}{2}\right).$$

Furthermore:

$$\phi_{+\infty}=\exp\left(K\frac{\pi}{2}\right)=V_2.$$

Similarly we obtain the coefficients c_n by an expansion of:

$$\phi(z)=\exp\left[K\left(-\frac{\pi}{2}+\arctan\frac{1}{z}\right)\right],$$

where

$$-\infty<z<-1,$$

i.e.,

$$\phi(z)=\phi_{-\infty}+c_0\cdot\frac{1}{z}+c_1\frac{1}{z^2}+c_2\cdot\frac{1}{z^3}+\cdots+c_2\cdot\frac{1}{z^{n+1}}+\cdots.$$

The relation between the a_n and the c_n is found to be:

$$c_n=a_{n+1}\exp\left(-K\frac{\pi}{2}\right),\quad \phi_{-\infty}=\exp\left(-K\frac{\pi}{2}\right)=V_1.$$

For the special voltage ratio $V_2/V_1=6$, K is equal to 0.571. The coefficients a_n then assume the following values:

$a_0=1$,	$a_5=0.0836$,	$a_{10}=0.0521$,
$a_1=0.571$,	$a_6=0.0775$,	$a_{11}=0.02821$,
$a_2=0.163$,	$a_7=0.0535$,	$a_{12}=0.04419$,
$a_3=0.1593$,	$a_8=0.0618$,	$a_{13}=0.02192$,
$a_4=0.10437$,	$a_9=0.0377$,	$a_{14}=0.03735$.

We did not attempt to prove generally that the solutions join on the surface of the sphere of radius 1. The actual calculation showed within the limits of computations with finite parts of infinite series that the solutions assume the same values on the respective parts of the sphere.

All series converge, though the convergence is better, of course, for values of r sufficiently far away from the value 1.

2. THE ELECTRODE CONFIGURATION

Due to the fact that the potential function is multivalued along the axis $z=0$ and $r>1$ it appears to be impossible to produce the axial potential distribution (6) by means of a simple electrode configuration. There is, however, a possibility of realizing the exact distribution in either one of the following two ways: (a) by means of an electrode configuration consisting of resistance material, (b) by the use of metallic rings arranged along the optical axis and charged to proper potentials. In both cases there is an infinite number of possibilities for the choice of the geometry of the arrangement, owing to the fact that any surface of rotational symmetry about the z axis can be chosen as basis. The axis along which the potential is double-valued must be excluded. According to method (a) one has to construct an electrode of resistive material whose inner surface is the chosen surface of rotation. If the circular ends of this surface are then connected to a potential source a current will flow along the walls of the surface. The driving potential and the resistance distribution can be chosen in such a way that a potential distribution is obtained which can be determined from the equipotential plot.[1] The method (b) also uses a surface of rotation as a basis, subdividing it into a number of rings whose radii are equal to the radius of the surface at the coordinate z. Each metallic ring is connected to a tap on a potential divider whose potential value can again be determined from the equipotential plot. In this manner the axial potential distribution could be achieved to any degree of accuracy. Figure 5 shows the potential distribution along a cylindrical surface of radius $r=1$, as well as the axial potential distribution.

3. THE SOLUTION OF THE PARAXIAL RAY EQUATION

While for most fields the solution of the paraxial can be achieved by approximate methods

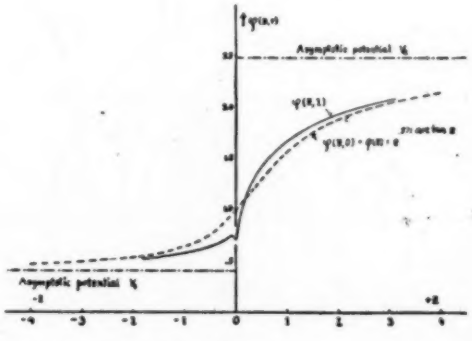

FIG. 5. Axial potential distribution and potential variation along a cylinder of radius $r=1$.

[1] I. G. Maloff and D. W. Epstein, *Electron Optics in Television* (McGraw-Hill Book Company, Inc., New York, 1938), first edition.

only, its solution for the special function:

$$\phi(z)=\phi_0 \exp\left(\frac{4}{3}\sqrt{3}k \arctan\frac{z}{a}\right) \tag{7}$$

can be given in general form in terms of elementary functions. It is:

$$\phi'(z)=\frac{4}{3}\sqrt{3}k\phi_0\frac{1}{a}\exp\left(\frac{4}{3}\sqrt{3}k \arctan\frac{z}{a}\right)\frac{1}{1+(z/a)^2};$$

hence:

$$\frac{3}{16}\left(\frac{\phi'(z)}{\phi(z)}\right)^2=\frac{(k/a)^2}{1+(z/a)^2}. \tag{8}$$

The paraxial ray differential equation for the case of a purely electrostatic lens is:[1-4]

$$\phi(z)r''(z)+\tfrac{1}{2}\phi'(z)r'(z)+\tfrac{1}{4}\phi''(z)r(z)=0; \tag{9}$$

and with:

$$r(z)=y(z)\cdot\exp\left(-\frac{1}{4}\int\frac{\phi'(u)}{\phi(u)}du\right)=y(z)\cdot\phi^{-\frac{1}{4}}(z),$$

Eq. (9) becomes:

$$y''(z)+\frac{3}{16}\left(\frac{\phi'(z)}{\phi(z)}\right)^2 y(z)=0. \tag{10}$$

Substituting (8) in Eq. (10) gives:

$$y''(z)+\frac{(k/a)^2}{[1+(z/a)^2]^2}y(z)=0. \tag{11}$$

Let:

$$\omega=(1+k^2)^{\frac{1}{2}}.$$

It has been shown by W. Glaser that the solution of (11) can be written in the form:

$$y(z)=\left(1+\left(\frac{z}{a}\right)^2\right)^{\frac{1}{4}}\left\{C_1\sin\left(\omega\,\mathrm{arccot}\frac{z}{a}\right)+C_2\cos\left(\omega\,\mathrm{arccot}\frac{z}{a}\right)\right\},$$

and because of:

$$\phi^{-\frac{1}{4}}(z)=\exp\left(-\tfrac{1}{3}\sqrt{3}k\arctan\frac{z}{a}\right),$$

$$r(z)=\left(1+\left(\frac{z}{a}\right)^2\right)^{\frac{1}{4}}\exp\left(-\tfrac{1}{3}\sqrt{3}k\arctan\frac{z}{a}\right)\cdot\{C_1\sin[\omega\,\mathrm{arccot}\,(z/a)]+C_2\cos[\omega\,\mathrm{arccot}\,(z/a)]\}.$$

This is the general solution of the paraxial ray equation describing the path of an electron through the lens whose axial potential is given by Eq. (7). The two constants of integration C_1 and C_2 (or K_1 and K_2 in the second form of the solution) are available to satisfy special conditions such as $r(z_1)=r_1$, $r'(z_1)=r_1'$ or $r(z_1)=r_1$, $r(z_2)=r_2$ for a given lens characterized by the two constants p and a.

It is useful to write the path equation in terms of the variable φ; where:

$$\cot\varphi=z/a.$$

Hence:

$$\tan((\pi/2)-\varphi)=\cot\varphi=z/a$$

or

$$\arctan(z/a)=(\pi/2)-\varphi,$$

and, therefore:

$$R(\varphi)=K_1\exp\left[\tfrac{1}{3}\sqrt{3}k\left(\varphi-\frac{\pi}{2}\right)\right]\cdot\frac{\sin\omega(\varphi+K_2)}{\sin\varphi}. \tag{12}$$

4. THE DETERMINATION OF THE CARDINAL POINTS; THE EXPRESSIONS FOR THE FOCAL LENGTHS

In order to determine the location of the focal points, the constant K_2 in Eq. (12) is chosen so that the path remains finite at $z=+\infty$ or $\varphi=0$. It is then:

$$R(\varphi)=K_1\exp\left[\tfrac{1}{3}\sqrt{3}k\left(\varphi-\frac{\pi}{2}\right)\right]\frac{\sin\omega\varphi}{\sin\varphi} \tag{13}$$

[2] E. Brueche and O. Scherzer, *Geometrische Elektronenoptik* (Verlagsbuchhandlung Julius Springer, Berlin, 1934).
[3] L. M. Myers, *Electron Optics* (Chapman and Hall, Ltd., London, 1939).
[4] V. K. Zworykin and G. A. Morton, *Television* (John Wiley and Sons, Inc., New York, 1940).

for K_2 must be equal to zero. The slope of this path is equal to zero at $z=+\infty$:

$$\frac{dr}{dz}=\frac{dR}{d\varphi}\cdot\frac{d\varphi}{dz}=K_1\left\{\frac{\sin\omega\varphi\cdot\cos\omega\varphi-\sin\omega\varphi\cdot\cos\varphi}{\sin^2\varphi}\exp\left[\tfrac{1}{3}\sqrt{3}k\left(\varphi-\frac{\pi}{2}\right)\right]\right.$$

$$\left.+\frac{\sin\omega\varphi}{\sin\varphi}\cdot\tfrac{1}{3}\sqrt{3}k\exp\left[\tfrac{1}{3}\sqrt{3}k\left(\varphi-\frac{\pi}{2}\right)\right]\right\}\left(-\frac{\sin^2\varphi}{a}\right).$$

Hence:

$$\left.\frac{dr(z)}{dz}\right|_{z=\infty}=0.$$

That is, the electron enters the lens parallel to the optical axis. Its distance from the axis is at that point:

$$r(+\infty)=K_1\omega\exp\left(-\tfrac{1}{3}\sqrt{3}k\frac{\pi}{2}\right). \tag{14}$$

For a ray entering the lens at $z=-\infty$ parallel to the axis the corresponding expressions are:

$$R(\varphi)=K_1\exp\left[\tfrac{1}{3}\sqrt{3}k\left(\varphi-\frac{\pi}{2}\right)\right]\frac{\sin[\omega(\varphi-\pi)]}{\sin\varphi},\quad \left.\frac{dr(z)}{dz}\right|_{z=-\infty}=0,\quad r(-\infty)=-K_1\omega\exp\left(\tfrac{1}{3}\sqrt{3}k\frac{\pi}{2}\right). \tag{15}$$

According to the definition of the focal points their location on the optical axis is given by:

$$\sin[\omega\varphi_{F_0}]=0$$

from Eq. (13). The index (o) shall indicate the object side of the lens, i.e., the side of the lens with negative abscissae. The other side is called the image side and is characterized by the index (i). The focal points on the object side are then:

$$\omega\varphi_{F_0}=n\cdot\pi,\quad n=1,2,3,\cdots\quad\text{or}\quad \varphi_{F_0}=n\cdot\pi/\omega,\quad n=1,2,3,\cdots.$$

$n=1$ is the first focal point through which the ray coming from $z=+\infty$ or $\varphi=0$ passes.

The focal points on the image side are determined by:

$$\sin[\omega(\varphi_{F_i}-\pi)]=0,$$

hence

$$\omega(\varphi_{F_i}-\pi)=-n\pi;\quad \varphi_{F_i}=\pi-n\pi/\omega;\quad n=1,2,3,\cdots$$

or in terms of the axial coordinate z:

$$z_{F_0}=a\cot(n\pi/\omega),\quad z_{F_i}=-a\cot(n\pi/\omega).$$

Knowing the location of the focal points, it is now possible to determine the corresponding focal lengths. These focal lengths are defined by:

$$f_0=\frac{r(+\infty)}{r'(z_{F_0})}=\text{focal length on the object side,}$$

$$f_i=\frac{r(-\infty)}{r'(z_{F_i})}=\text{focal length on the image side.}$$

Using Eqs. (14) and (15) and the following expressions:

$$r'(z_{F_0})=K_1\sin n\frac{\pi}{\omega}\exp\left[\tfrac{1}{3}\sqrt{3}k\left(n\frac{\pi}{\omega}-\frac{\pi}{2}\right)\right]\cdot\omega\cdot(-1)^{n-1}\frac{1}{a},$$

$$r'(z_{F_i})=K_1\sin\left(\pi-n\frac{\pi}{\omega}\right)\exp\left[\tfrac{1}{3}\sqrt{3}k\left(\pi-n\frac{\pi}{\omega}-\frac{\pi}{2}\right)\right]\cdot\omega\cdot(-1)^n\cdot\left(-\frac{1}{a}\right),$$

f_0 and f_i become:

$$f_0=(-1)^{n-1}\frac{a\exp\left(-\frac{1}{3}\sqrt{3}kn\frac{\pi}{\omega}\right)}{\sin n\frac{\pi}{\omega}},$$

$$f_i=(-1)^{n}\frac{a\exp\left(+\frac{1}{3}\sqrt{3}kn\frac{\pi}{\omega}\right)}{\sin n\frac{\pi}{\omega}}. \tag{16}$$

There will be a value of k for which f_0/a assumes a minimum. This value of k will, however, not give a minimum for f_i/a. In case of the corresponding magnetic lens investigated by W. Glaser[5] the two focal lengths are equal and assume their minimum value therefore for the same value of the lens parameter.

The condition for the minimum for f_0/a is:

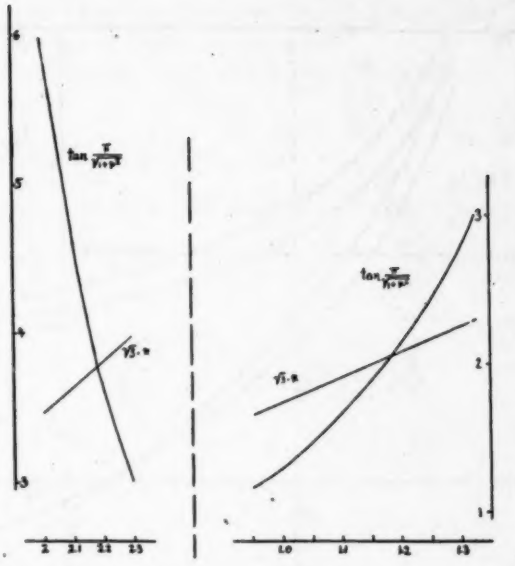

FIG. 6. Graphs for the determination of the minimum values of the focal lengths.

$$[d(f_0/a)]/dk=0,$$

which yields the transcendental equation:

$$\tan \pi/(1+k^2)^{\frac{1}{2}}=\sqrt{3}k. \tag{17}$$

For the other focal length f_i/a the condition:

$$[d(f_i/a)]/dk=0,$$

yields:

$$\tan \pi/(1+k^2)^{\frac{1}{2}}=-\sqrt{3}k. \tag{18}$$

The two values of k satisfying the two Eqs. (17) and (18) can be obtained graphically. These values are:

$$k_{fi}=1.185 \quad \text{or} \quad (V_2/V_1)_{fi}=5404,$$

$$k_{f0}=2.178 \quad \text{or} \quad (V_2/V_1)_{f0}\cong 7{,}275{,}000$$

(see Fig. 6).

If the voltage ratio V_2/V_1 increases, the focal length on the image side assumes its minimum before the focal length on the object side does. These minimum values are:

$$f_0/a=\ \ 0.199$$
$$f_i/a=-4.45.$$

The values of V_2/V_1 are very high so that it seems almost impossible to obtain the minimum focal lengths in practice.

The focal lengths and the location of the focal points as a function of the voltage ratio are plotted in Figs. 7 and 8.

[5] W. Glaser, Zeits. f. Physik **117**, 285 (1941).

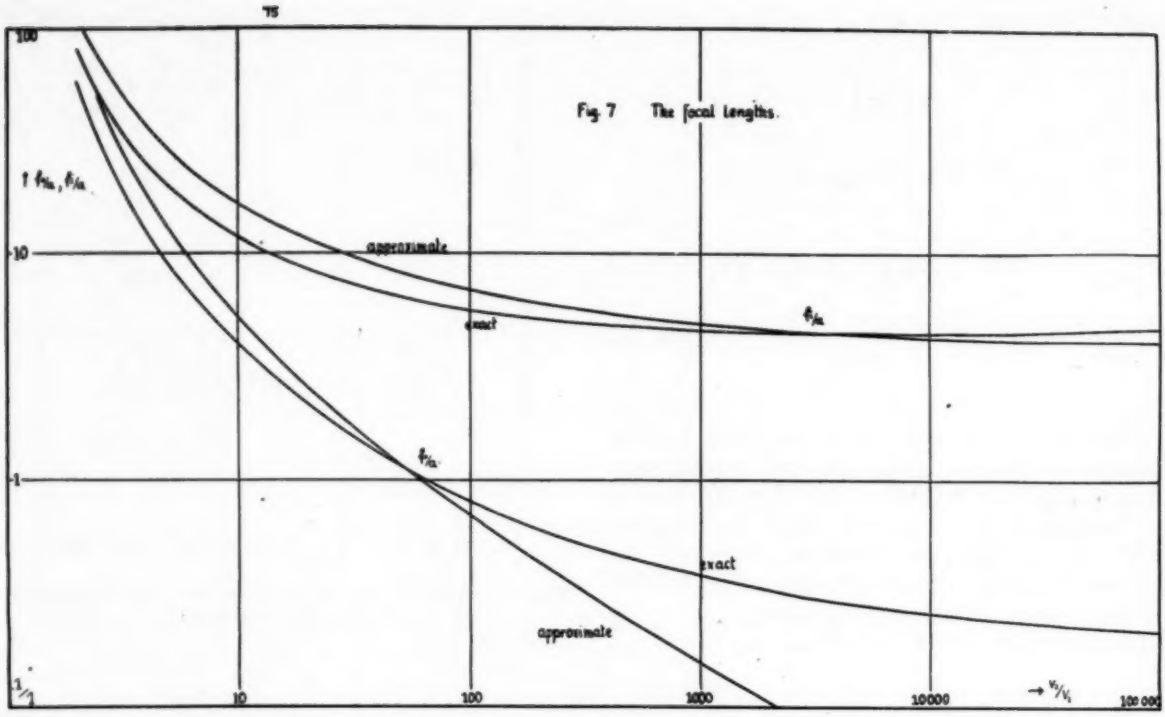

FIG. 7. The focal lengths.

5. COMPARISON WITH DESIGN FORMULAE FOR THIN LENSES

Due to the fact that it is impossible to obtain accurate formulae for the focal lengths of electron lenses approximate formulae are often used for design purposes. It is therefore worth while to compare these formulae with the accurate ones we obtained for the special lens discussed here.

These approximate formulae for the focal lengths are:

$$\frac{1}{f_0}=\frac{1}{8}\frac{1}{[\phi(+\infty)]^{\frac{1}{2}}}\int_{-\infty}^{+\infty}\frac{[\phi'(z)]^2}{\phi^{\frac{1}{2}}(z)}dz, \tag{19}$$

$$\frac{1}{f_i}=-\frac{1}{8}\frac{1}{[\phi(-\infty)]^{\frac{1}{2}}}\int_{-\infty}^{+\infty}\frac{[\phi'(z)]^2}{\phi^{\frac{1}{2}}(z)}dz. \tag{20}$$

With

$$\phi(z)=\exp\left[\frac{4}{3}\sqrt{3}k\arctan\left(\frac{z}{a}\right)\right]=\exp\left[\frac{4}{3}\sqrt{3}k\left(\frac{\pi}{2}-\varphi\right)\right],$$

$$\phi'(z)=\frac{4}{3}\sqrt{3}k\frac{1}{a}\exp\left[\frac{4}{3}\sqrt{3}k\arctan\left(\frac{z}{a}\right)\right]\cdot\frac{1}{1+(z/a)^2}=\frac{4}{3}\sqrt{3}\frac{k}{a}\exp\left[\frac{4}{3}\sqrt{3}k\left(\frac{\pi}{2}-\varphi\right)\right]\cdot\sin^2\varphi,$$

$$dz=-\frac{a}{\sin^2\varphi}d\varphi$$

the expressions (19) and (20) take the form:

$$\frac{f_0}{a}=\frac{2(k^2+3)}{\sqrt{3}k}\cdot\frac{1}{\exp\left(\frac{2}{3}\sqrt{3}k\pi\right)-1},\quad \frac{f_i}{a}=\frac{-2(k^2+3)}{\sqrt{3}k}\cdot\frac{1}{1-\exp\left(-\frac{2}{3}\sqrt{3}k\pi\right)}. \tag{21}$$

The focal lengths are represented graphically in Fig. 7. It can be seen from these graphs that for a fairly wide range of voltage ratios the formulae for thin lenses give a relatively good approximation.

This is a remarkable fact considering that the comparison between the corresponding formulae for the magnetic lens proved to be very unfavorable for the approximate expressions.

Another expression in the theory of thin lenses relates to the ratio of the focal lengths. It is:[4]

$$\frac{f_i}{f_0}=\left[\frac{\phi(+\infty)}{\phi(-\infty)}\right]^{\frac{1}{2}}. \tag{22}$$

From the approximate expressions (21) this ratio is indeed given as the square root of the ratio of the asymptotic voltages:

$$\frac{f_i}{f_0}=\exp\left(\tfrac{2}{3}\sqrt{3}k\pi\right).$$

Our accurate formulae on the other hand yield:

$$\frac{f_i}{f_0}=\exp\left[\tfrac{2}{3}\sqrt{3}k\frac{\pi}{(1+k^2)^{\frac{1}{2}}}\right].$$

Both expressions are approximately equal for small values of k, as can be seen from the equations or from Fig. 9, where the ratio of the focal lengths is plotted as a function of the voltage ratio.

6. NEWTON'S IMAGE RELATIONS

In the preceding paper[6] we have shown that Newton's image relations hold for the special lens investigated here. This relation, however, can be derived independently of these considerations. It is known that object and image coordinates for any lens are related by:

$$r_1(z_0)\cdot r_2(z_i)-r_1(z_i)\cdot r_2(z_0)=0, \tag{23}$$

where $r_1(z)$ and $r_2(z)$ are any two linearly independent solutions of the paraxial ray equation.[5]

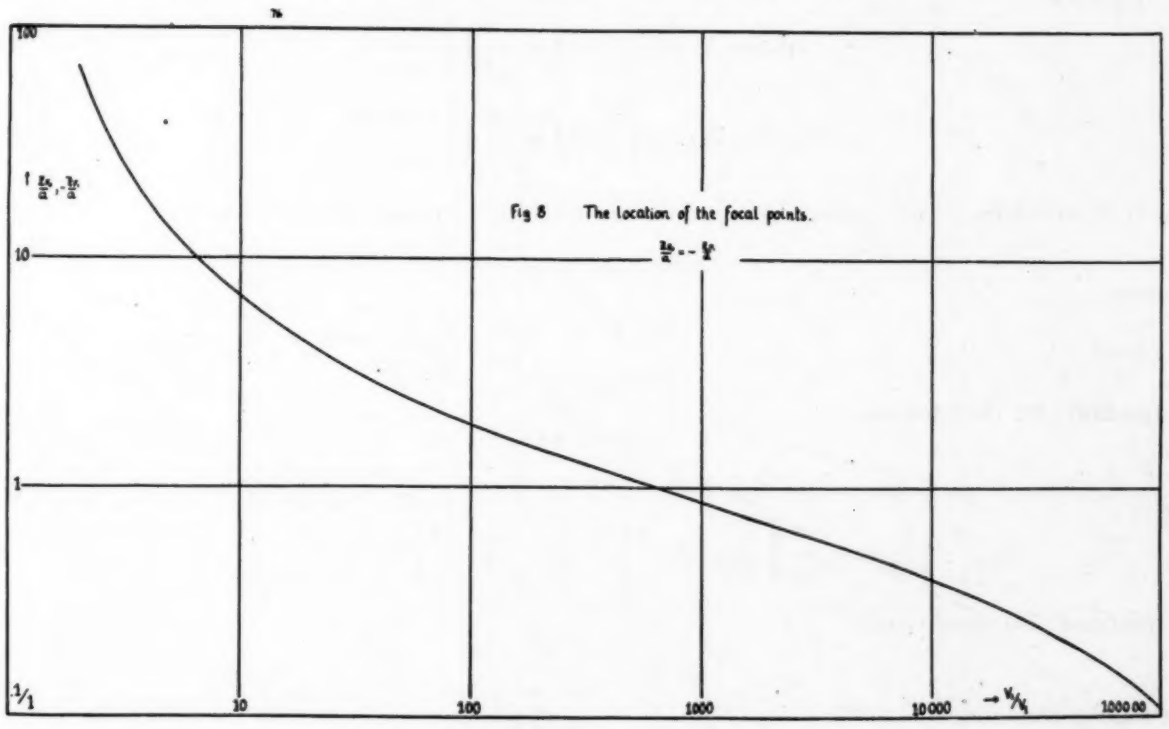

FIG. 8. The location of the focal points.

[6] R. G. E. Hutter, J. App. Phys. 16, 670 (1945).

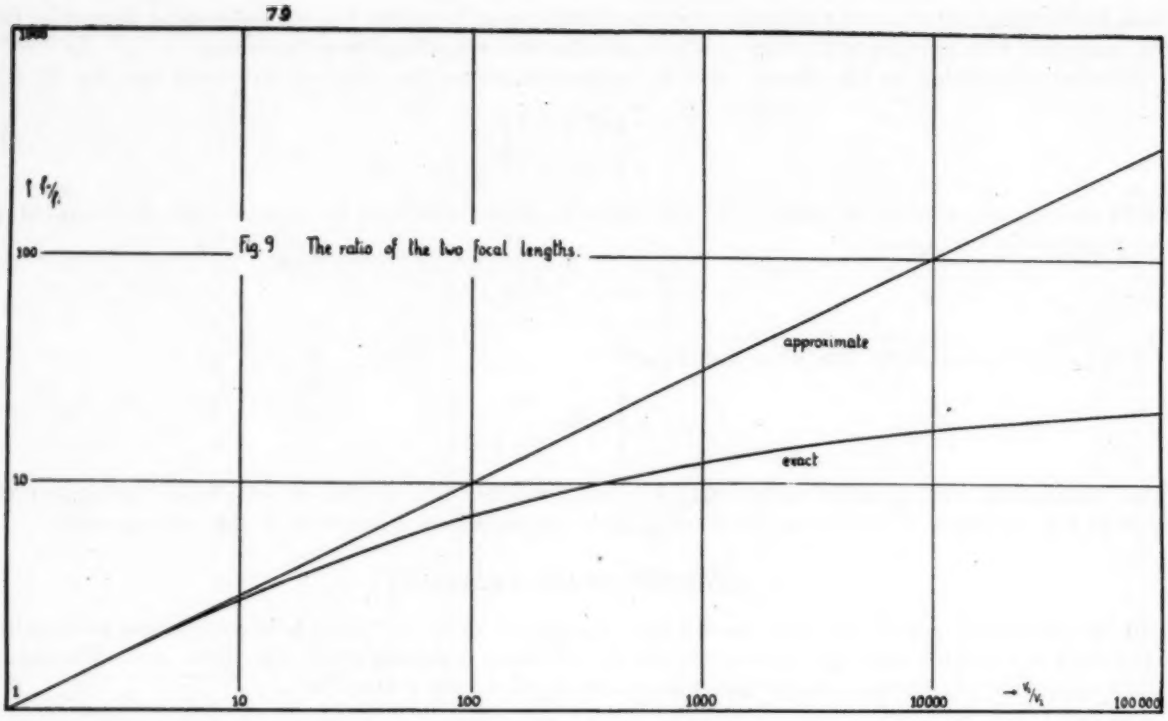

Fig. 9 The ratio of the two focal lengths.

FIG. 9. The ratio of the two focal lengths.

We take

$$r_1(\varphi)=K_1\exp\left[\tfrac{1}{3}\sqrt{3}k\left(\varphi-\frac{\pi}{2}\right)\right]\cdot\frac{\sin\omega\varphi}{\sin\varphi}, \tag{24}$$

$$r_2(\varphi)=K_2\exp\left[\tfrac{1}{3}\sqrt{3}k\left(\varphi-\frac{\pi}{2}\right)\right]\cdot\frac{\cos\omega\varphi}{\sin\varphi} \tag{25}$$

as such solutions. Substituting (24) and (25) in Eq. (23), this equation becomes:

$$\sin\left[\omega(\varphi_0-\varphi_i)\right]=0,$$

hence:

$$\varphi_0-\varphi_i=n\cdot\pi/\omega. \tag{26}$$

It was:

$$\cot\varphi=z/a.$$

Equation (26) then becomes:

$$\frac{z_0}{a}=\frac{z_i/a\cot n\pi/\omega-1}{z_i/a+\cot n\pi/\omega},$$

or:

$$\left[\frac{z_0}{a}-\cot n\frac{\pi}{\omega}\right]\left[\frac{z_i}{a}+\cot n\frac{\pi}{\omega}\right]=-\left[1+\cot^2 n\frac{\pi}{\omega}\right]=-\frac{1}{\sin^2 n\pi/\omega}.$$

Equations (16) show that:

$$f_0\cdot f_i=-\frac{a^2}{\sin^2 n\pi/\omega}.$$

It follows therefore:

$$(z_0-z_{F_0})(z_i-z_{F_i})=f_0\cdot f_i \tag{27}$$

which is Newton's image relation.

In order to derive the magnification formula a general formula for the magnification of any electron lens is used as the starting point. If $r(z)$ is any particular solution of the paraxial ray equation, the magnification can be expressed by:[5]

$$M=r(z_i)/r(z_0). \tag{28}$$

If we take the particular solution given by Eq. (25) the expression (28) becomes:

$$M=\exp\left[\tfrac{1}{3}\sqrt{3}k(\varphi_i-\varphi_0)\right]\cdot\frac{\cos(\omega\varphi_i)}{\sin(\omega\varphi_i)}\cdot\frac{\sin\varphi_0}{\cos(\omega\varphi_0)}=\exp\left(-\tfrac{1}{3}\sqrt{3}kn\pi/\omega\right)\cdot\frac{\cos(\omega\varphi_0-n\pi)}{\sin(\varphi_0-n\pi)}\cdot\frac{\sin\varphi_0}{\cos(\omega\varphi_0)}$$

$$=(-1)^n\frac{\exp\left(-\tfrac{1}{3}\sqrt{3}kn\pi/\omega\right)}{\cot n\pi/\omega-\cot\varphi_0}\cdot\frac{1}{\sin n\pi/\omega}=(-1)^n(-1)^{n-1}\frac{f_0}{a}\cdot\frac{1}{z_{F_0}/a-z_0/a}$$

hence:

$$M=\frac{f_0}{z_0-z_{F_0}}=\frac{z_i-z_{F_i}}{f_i}. \tag{29}$$

The simple geometrical image construction based on the two expressions (27) and (29) is illustrated for a special choice of the voltage ratio in Fig. 10. The two principal rays used have nothing in common with the actual electron path except the end points. Within the limits of the Gaussian dioptrics (i.e., neglecting aberrations) the influence of the field can be replaced by the two principal planes and the focal points.

The principal planes are those planes with unit magnification. With $M=1$ in Eq. (24) these planes have the following axial coordinates:

$$z_{H_0}=f_0+z_{F_0}=a\frac{(-1)^{n-1}\exp\left(-\tfrac{1}{3}\sqrt{3}kn\pi/\omega\right)+\cos(n\pi/\omega)}{\sin(n\pi/\omega)},$$

$$z_{H_i}=f_i+z_{F_i}=a\frac{(-1)^n\exp\left(+\tfrac{1}{3}\sqrt{3}kn\pi/\omega\right)-\cos(n\pi/\omega)}{\sin(n\pi/\omega)}.$$

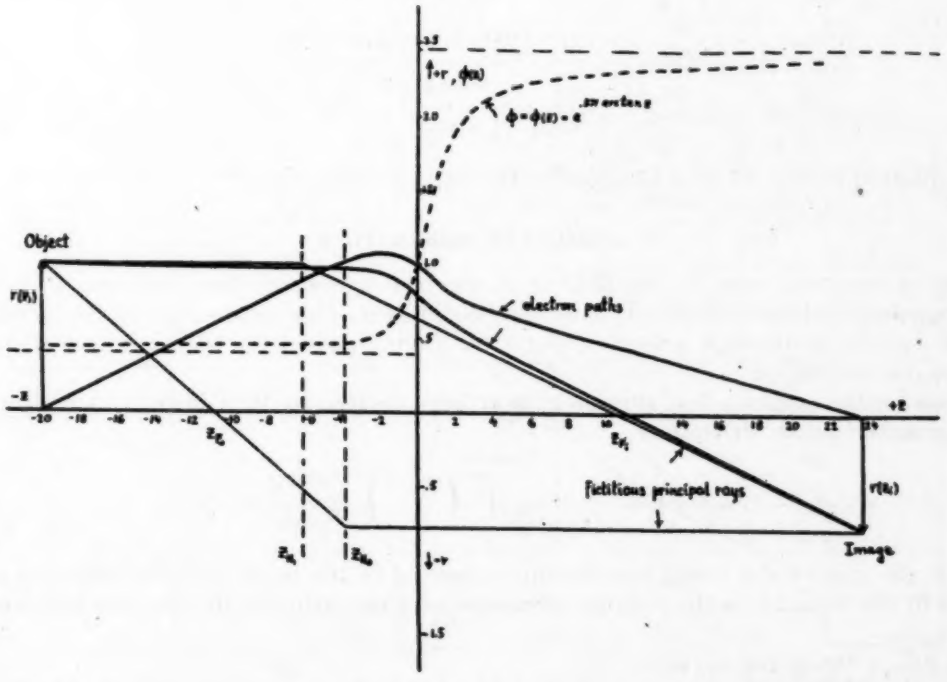

FIG. 10. Image construction for the special lens with the axial potential distribution $\phi(z)e^{-\sqrt{3}\arctan z}$.

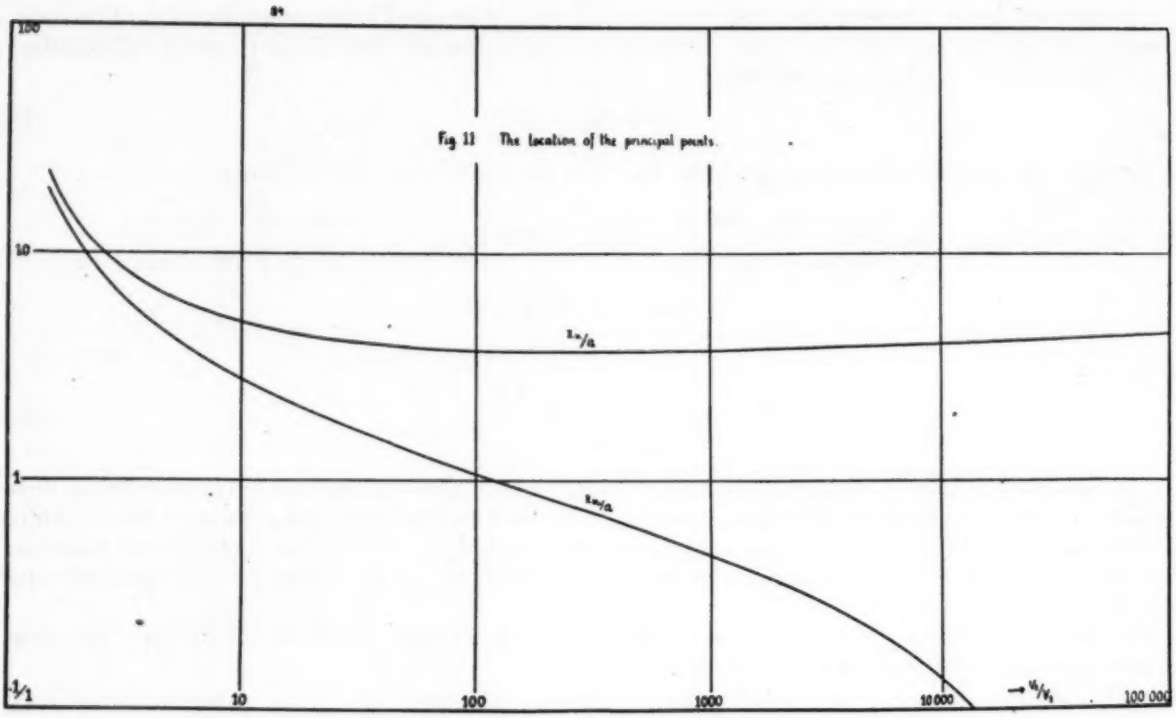

FIG. 11. The location of the principal points.

If $n=1$

$$z_{H_0}=a\frac{\exp\left(-\frac{1}{3}\sqrt{3}k\cdot\pi/\omega\right)+\cos\left(\pi/\omega\right)}{\sin\left(\pi/\omega\right)},$$

$$z_{H_i}=-a\frac{\exp\left(\frac{1}{3}\sqrt{3}k\cdot\pi/\omega\right)-\cos\left(\pi/\omega\right)}{\sin\left(\pi/\omega\right)},$$

where

$$\omega=(1+k^2)^{\frac{1}{2}}.$$

z_{H_0}, z_{H_i} are plotted in Fig. 11 as a function of the voltage ratio V_2/V_1.

7. CHROMATIC ABERRATION

If a beam of electrons entering the field of an electron lens is not homogeneous with regard to speed the curvatures of the individual paths will be different. This means that electrons leaving one point will not again go through a single point after having passed the focusing field. This effect is called chromatic aberration.

Expressions for the magnitude of this effect have been derived by W. Glaser[7] and A. Recknagel.[8] Glaser's expression can be written as:

$$\delta_{\text{chr}}=\tfrac{3}{8}\alpha\cdot\Delta V\frac{1}{[\phi(z_0)]^{\frac{1}{2}}}\int_{z_0}^{z_i}\left(\frac{\phi'(z)}{\phi(z)}\right)^2\cdot\frac{r^2(z)}{[\phi(z)]^{\frac{1}{2}}}dz. \tag{30}$$

where $\phi(z)$ is the sum of the initial accelerating potential of the beam and the potential up to the coordinate z in the lens. ΔV is the voltage inhomogeneity occurring in the electron beam satisfying

[7] W. Glaser, Zeits. f. Physik **116**, 56 (1940).
[8] A. Recknagel, *AEG Forschungsinstitut, Jahrbuch. Sonderheft: Uebermikroskop* (Verlagsbuchhandlung Julius Springer, Berlin, 1940).

the condition: $\Delta V \ll \phi(z)$. $r=r(z)$ is a solution of Eq. (9) which satisfies the following conditions:

$$r(z_0)=r(z_i)=0; \quad r'(z_0)=1. \tag{31}$$

α is the angular aperture of the electron beam at the object position.

For the case of the electrostatic lens discussed here the expression can be evaluated. We obtain:

$$\delta_{\text{chr}}=\alpha\Delta V\cdot\frac{3}{4}\frac{\sqrt{3}k\cdot a}{7k^2+3}\cdot\frac{\exp\left[\frac{4}{3}\sqrt{3}k\left(\varphi_0-\frac{\pi}{2}\right)\right]}{\sin^2\varphi_0}\left\{1-\exp\left[-\frac{4}{3}\sqrt{3}k\frac{\pi}{(1+k^2)^{\frac{1}{2}}}\right]\right\}. \tag{32}$$

The dimensionless quantity:

$$\frac{C_{\text{chr}}}{a}=\frac{3}{4}\frac{\sqrt{3}k}{7k^2+3}\frac{\exp\left[\frac{4}{3}\sqrt{3}k\left(\varphi_0-\frac{\pi}{2}\right)\right]}{\sin^2\varphi_0}\left\{1-\exp\left[-\frac{4}{3}\sqrt{3}k\frac{\pi}{(1+k^2)^{\frac{1}{2}}}\right]\right\}$$

characterizes the property of the lens as regards chromatic aberration. C_{chr}/a is a function of k as well as of the object position φ_0. We will show that an optimum position $\varphi_{0\,\text{opt}}$ exists so that C_{chr}/a is a minimum. From:

$$\frac{d(C_{\text{chr}}/a)}{d\varphi_0}=0,$$

it follows that:

$$\frac{4}{3}\sqrt{3}k\sin\phi_0 = 2\cos\varphi_0;$$

or

$$\cot\varphi_{0\,\text{opt}}=\frac{z_{0\,\text{opt}}}{a}=\tfrac{2}{3}\sqrt{3}k.$$

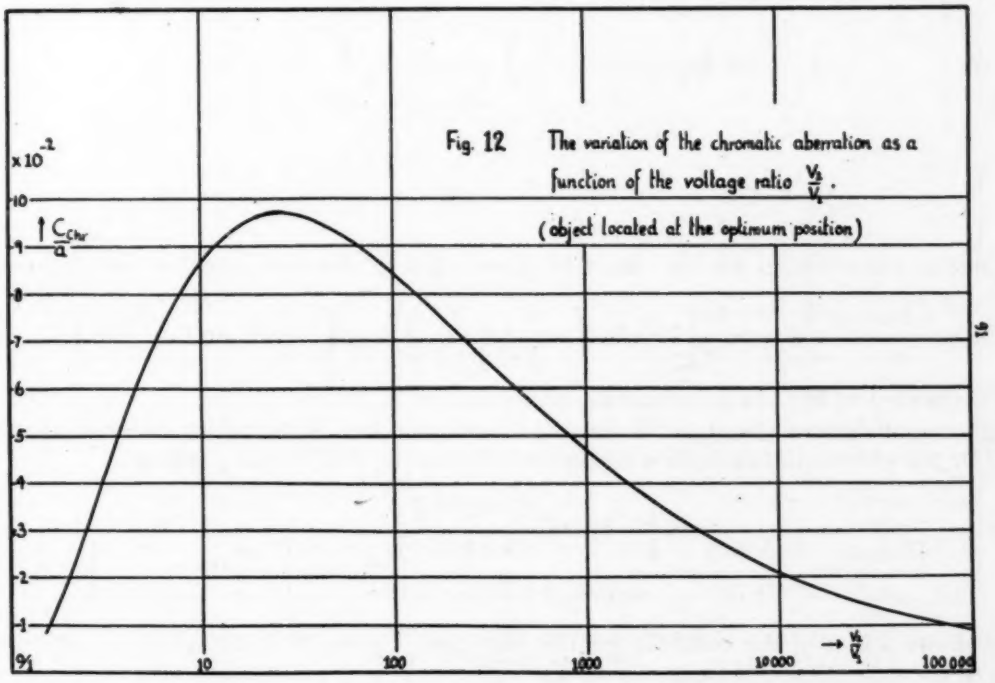

FIG. 12. The variation of the chromatic aberration as a function of the voltage ratio V_2/V_1 (object located at the optimum position).

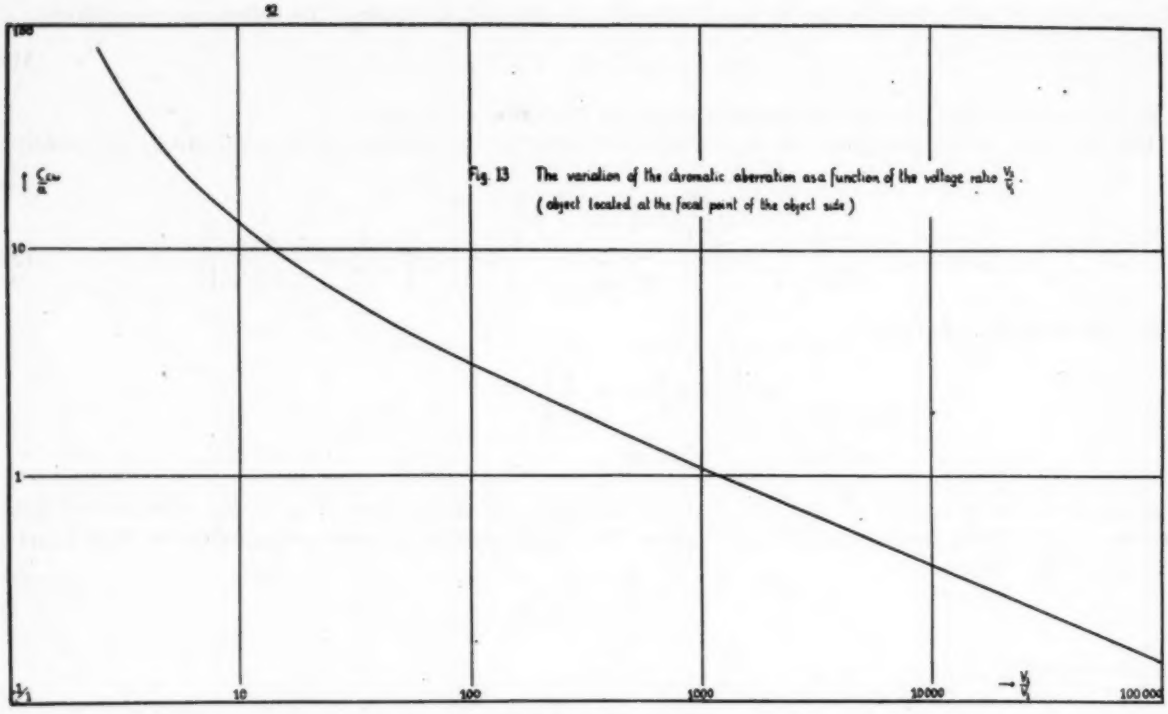

FIG. 13. The variation of the chromatic aberration as a function of the voltage ratio V_2/V_1 (object located at the focal point of the object side).

It is interesting to note that this point is the same as that for which the axial potential distribution has its point of inflection. For such a point the second derivative must be equal to zero, i.e.,

$$\frac{d^2 \exp\left(\frac{4}{3}\sqrt{3}k \arctan\frac{z_0}{a}\right)}{d(z_0/a)^2}=0, \quad \text{or} \quad \frac{4}{3}\sqrt{3}k-2\frac{z_0}{a}=0.$$

Hence:

$$\frac{z_{0\ \text{infl}}}{a}=\tfrac{2}{3}\sqrt{3}k.$$

The value of the constant for the chromatic aberration at this optimum position is given by:

$$\frac{C_{\text{chr opt}}}{a}=\frac{\sqrt{3}}{4}k\cdot\frac{3+4k^2}{3+7k^2}\left[1-\exp\left(-\frac{4}{3}\sqrt{3}k\frac{\pi}{1+k^2}\right)\right]\exp\left(-\frac{4}{3}\sqrt{3}k \arctan\frac{2}{3}\sqrt{3}k\right).$$

$C_{\text{chr opt}}/a$ is plotted in Fig. 12 as a function of V_2/V_1.

For high magnifications the object is near the focal point zF_0. Substituting the value for $\varphi_0=\pi/\omega$ in Eq. (32), the chromatic aberration constant becomes for this object position:

$$\frac{C_{\text{chr focal}}}{a}=\frac{3}{4}\frac{\sqrt{3}k}{3+7k^2}\frac{\exp\left[-\frac{2}{3}\sqrt{3}k\pi\frac{(1+k^2)^{\frac{1}{2}}-2}{(1+k^2)^{\frac{1}{2}}}\right]}{\sin^2\pi/(1+k^2)^{\frac{1}{2}}}\left\{1-\exp\left[-\frac{4}{3}\sqrt{3}k\frac{\pi}{(1+k^2)^{\frac{1}{2}}}\right]\right\}.$$

Figure 13 shows a plot of the constant for this last case. Figure 14 finally gives a graphical representation of the dependence of $(z_{0\ \text{opt}}/a)$ on V_2/V_1.

The dependence of the chromatic aberration constant on the position of the object is shown in Fig. 15 for the special case $V_2/V_1=6$.

8. SPHERICAL ABERRATION

It has been pointed out before that perfect imagery could only be obtained if the rays through an electron lens would satisfy a differential equation of the type of the paraxial ray equation.[9] Since this is the case approximately only for the so-called "paraxial" rays the imagery of real electron lenses is never perfect. If the angular aperture is increased the image of a point becomes a small disk whose radius increases with the third power of the angular aperture. This radius divided by the magnification M serves as a measure of the aperture defect termed "spherical aberration."[2]

Glaser and Scherzer[10] have derived expressions for:

$$\delta_{sp}=\Delta r/M$$

in terms of the field function along the optical axis:

$$\phi=\phi(z), \quad H=H(z)$$

and the paraxial ray path function:

$$r=r(z).$$

For a purely electrostatic lens Scherzer's expression becomes:

$$\delta_{sp}=\alpha^3\frac{1}{16[\phi(z_0)]^{\frac{1}{2}}}\int_{z_0}^{z_i}\phi^{-\frac{1}{2}}\cdot\left[\frac{5}{4}\phi''^2+\frac{5}{24}\frac{\phi'^4}{\phi^2}+\frac{14}{3}\frac{\phi'^2}{\phi}\frac{r'}{r}-\frac{3}{2}\phi'^2\frac{r'^2}{r^2}\right]r^4dz. \tag{33}$$

The function $r(z)$ again satisfies the conditions (31):

$$r(z_0)=r(z_i)=0; \quad r'(z_0)=1 \tag{31}'$$

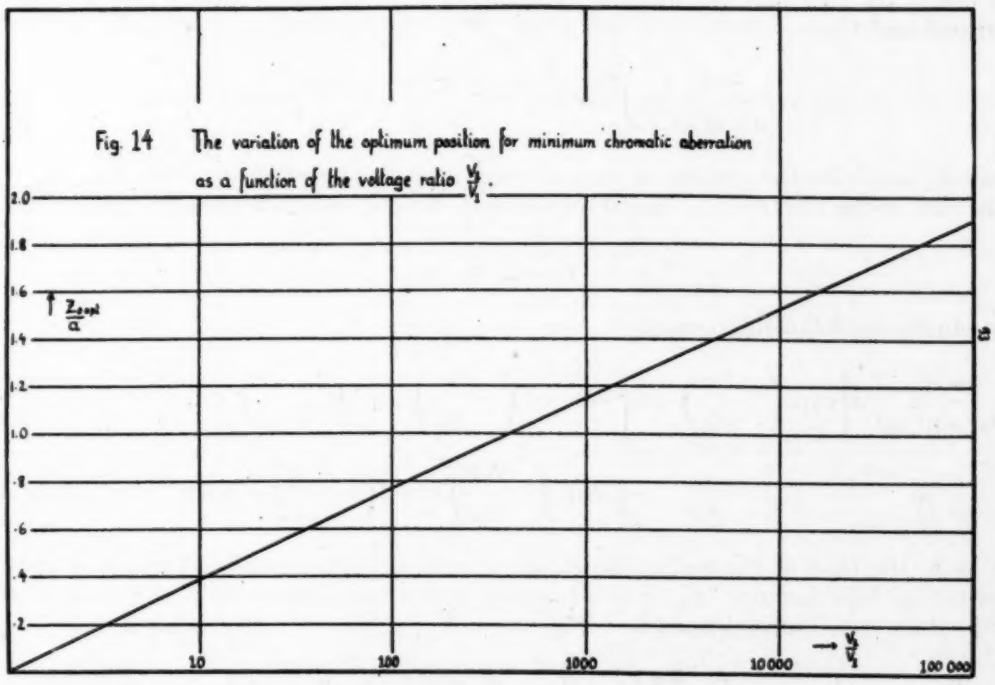

Fig. 14 The variation of the optimum position for minimum chromatic aberration as a function of the voltage ratio $\frac{V_2}{V_1}$.

FIG. 14. The variation of the optimum position for minimum chromatic aberration as a function of the voltage ratio V_2/V_1.

[9] L. Marton and R. G. E. Hutter, Proc. I.R.E. **32**, 3 (1944); Proc. I.R.E. **32**, 546 (1944).
[10] H. Busch and E. Brueche, *Beitraege zur Elektronenoptik* (Johann Ambrosius Barth, Leipzig, 1937).

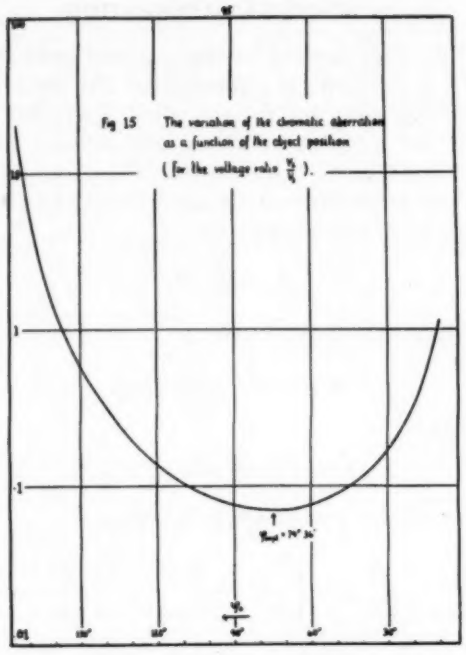

FIG. 15. The variation of the chromatic aberration as a function of the object position (for the voltage ratio V_2/V_1).

Making use of the fact that the function $r(z)$ vanishes at the end points of the integral, Eq. (33) can be transformed into:

$$\delta_{sp}=\alpha^3\frac{1}{16[\phi(z_0)]^{\frac{1}{2}}}\int_{z_0}^{z_i}\phi^{-\frac{1}{2}}\left\{\phi''^2-\frac{5}{2}\frac{\phi'^2\cdot\phi''}{\phi}+\frac{5}{2}\frac{\phi'^4}{\phi^2}-\frac{1}{4}\phi'\phi'''\right\}r^4dz \tag{34}$$

by repeatedly applying the process of partial integration and substitution from Eq. (9).

For the case of the electrostatic lens discussed here we can evaluate the expression (34). Let:

$$K=(4/3)\sqrt{3}k. \tag{35}$$

We then obtain the following formula:

$$\delta_{sp}=\frac{-a\alpha^3K^2}{512\omega^4\sin^4\varphi_0}\left\{A\left[\exp\left(-\frac{K\pi}{2\omega}\right)-1\right]-B\exp\left(-\frac{K\pi}{2\omega}\right)\sin 2\left(\varphi_0-\frac{\pi}{\omega}\right)\right.$$
$$\left.-C\exp\left(-\frac{K\pi}{2\omega}\right)\cos 2\left(\varphi_0-\frac{\pi}{\omega}\right)+B\sin 2\varphi_0+C\cos 2\varphi_0\right\}. \tag{36}$$

Again as in the case of chromatic aberration a constant can be defined which is characteristic of the particular field function. C_{sp}/a is a function of the lens parameter k as well as of the object position φ_0. This constant is defined by:

$$\frac{C_{sp}}{a}=\frac{K^2}{512\omega^2\sin^4\varphi_0}\left\{A\left[1-\exp\left(-\frac{K\pi}{2\omega}\right)\right]+B\left[\exp\left(-\frac{K\pi}{2\omega}\right)\sin 2\left(\varphi_0-\frac{\pi}{\omega}\right)-\sin 2\varphi_0\right]\right.$$
$$\left.+C\left[\exp\left(-\frac{K\pi}{2\omega}\right)\cos 2\left(\varphi_0-\frac{\pi}{\omega}\right)-\cos 2\varphi_0\right]\right\}, \tag{37}$$

where the coefficients A, B, C of Eqs. (36) and (37) are:

$$A=\frac{9225}{1024}\cdot\frac{1}{K}\cdot\frac{(16+3K^2)^2}{(64+13K^2)},$$

$$B=2\left\{\frac{11K^2+80}{K^2+16}-\frac{131K^4+512K^2-768}{169K^4+1312K^2+2304}\right\},$$

$$C=K\left\{\frac{3(3K^2+32)}{K^2+16}-\frac{3K^4-64}{K^2(K^2+4)}+\frac{39K^4-316K^2-2560}{169K^4+1312K^2+2304}\right\}.$$

Equation (37) can also be written as:

$$\frac{C_{sp}}{a}=\frac{K^2}{512\omega^4\sin^4\varphi_0}\{L+M\sin 2\varphi_0+N\cos 2\varphi_0\}, \tag{38}$$

where the coefficients L, M, and N are related to the A, B, and C by the following equations:

$$L=A\left[1-\exp\left(-\frac{K\pi}{2\omega}\right)\right],$$

$$M=B\left[\exp\left(-\frac{K\pi}{2\omega}\right)\cos\frac{2\pi}{\omega}-1\right]+C\exp\left(-\frac{K\pi}{2\omega}\right)\sin\frac{2\pi}{\omega},$$

$$N=C\left[\exp\left(-\frac{K\pi}{2\omega}\right)\cos\frac{2\pi}{\omega}-1\right]-B\exp\left(-\frac{K\pi}{2\omega}\right)\sin\frac{2\pi}{\omega}.$$

Just as was the case for the chromatic aberration constant there exists an optimum position φ_m or

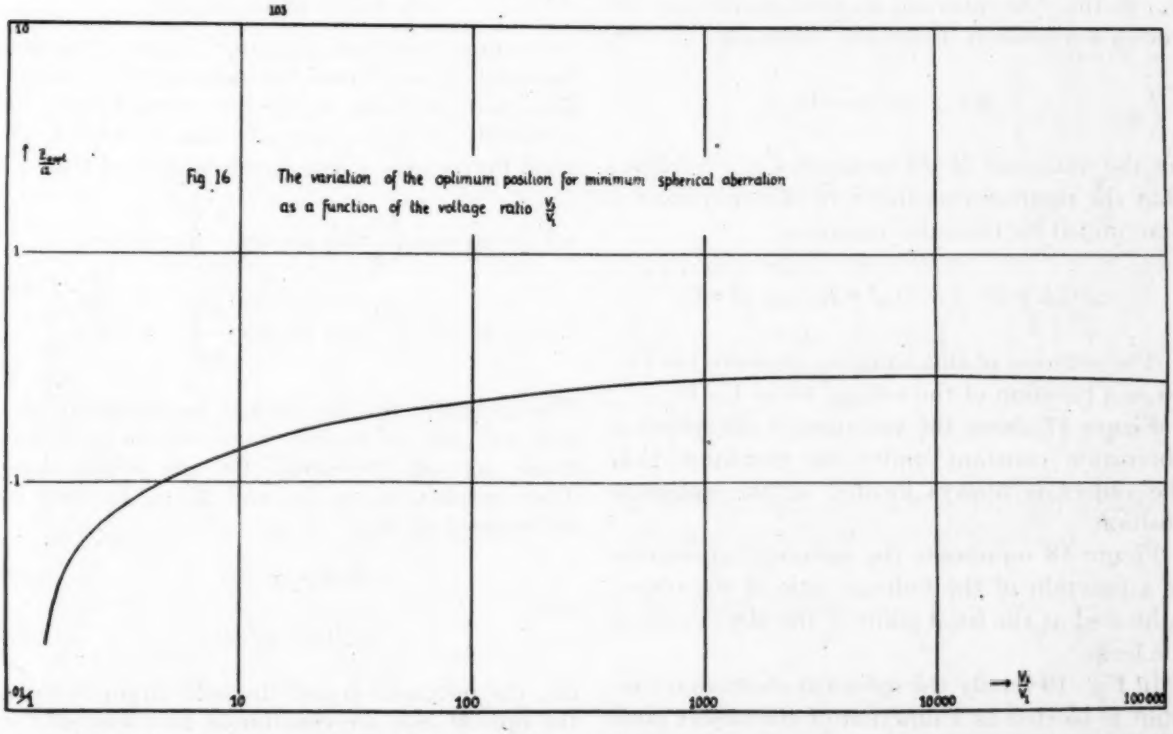

Fig. 16 The variation of the optimum position for minimum spherical aberration as a function of the voltage ratio $\frac{V_2}{V_1}$

FIG. 16. The variation of the optimum position for minimum spherical aberration as a function of the voltage ratio V_2/V_1.

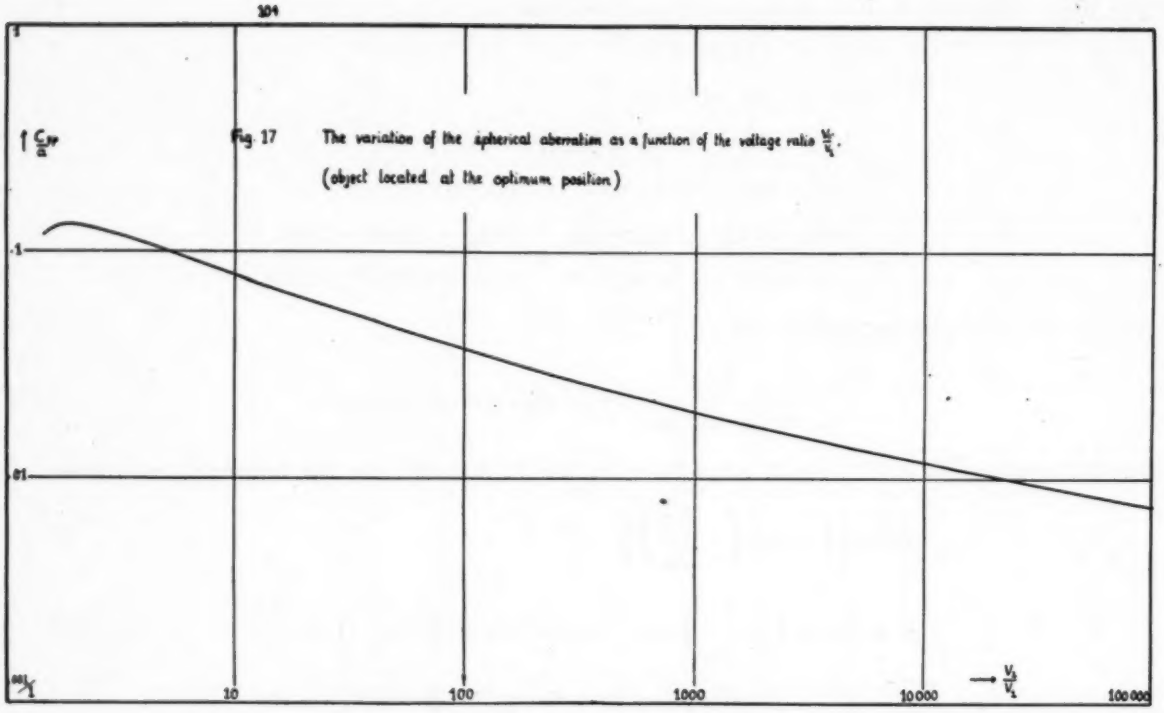

FIG. 17. The variation of the spherical aberration as a function of the voltage ratio V_2/V_1 (object located at the optimum position).

x_m, so that the spherical aberration constant becomes a minimum. From the condition:

$$d(C_{sp}/a)/dx_0=0,$$

for the minimum of the constant C_{sp}/a it follows that the position coordinate of this minimum is determined by the cubic equation:

$$x_m^3(2L+M)+3Mx_m^2+2Lx_m+M=0.$$

The solution of this equation is plotted in Fig. 16 as a function of the voltage ratio V_2/V_1.

Figure 17 shows the variation of the spherical aberration constant under the condition that the object is always located at the optimum position.

Figure 18 represents the spherical aberration as a function of the voltage ratio if the object is located at the focal point of the object side of the lens.

In Fig. 19 finally the spherical aberration constant is plotted as a function of the object position for the special voltage ratio of $V_2/V_1=6$.

9. A GENERALIZATION

In order to obtain a greater degree of freedom in matching axial potential distributions of simple lens configurations to the one treated here theoretically, a slight generalization is helpful. We shall investigate a lens which is defined by:

$$\phi_1(z)=\phi_{01}\exp\left(\frac{4}{3}\sqrt{3}k_1\arctan\frac{z}{a_1}\right), \quad z\leqq 0$$
$$\phi_2(z)=\phi_{02}\exp\left(\frac{4}{3}\sqrt{3}k_2\arctan\frac{z}{a_2}\right). \quad z\geqq 0 \qquad (38a)$$

The formulae for the optical quantities of this lens will only be slightly more complicated than those derived previously for the simple lens. The constants ϕ_{01} a_1, k_1, and ϕ_{02} a_2, k_2 must be determined so that:

$$\phi_1(0)=\phi_2(0), \qquad (39)$$

and

$$\phi_1'(0)=\phi_2'(0), \qquad (40)$$

i.e., the potential ϕ and the field strength along the optical axis are continuous functions of the axial coordinate z.

Equation (39) requires that:

$$\phi_{01}=\phi_{02}$$

and from Eq. (40) it follows that:

$$k_1/a_1=k_2/a_2.$$

If only *one* expression of the form (1) is used to describe a potential distribution which has two asymptotic potentials V_2 and V_1 the constants k and ϕ_0 are completely determined. If, on the other hand, *two* expressions of this form are chosen, k_1 and k_2 are only determined after a convenient choice of ϕ_0 has been made. a_1 and a_2 are again related to the slope of the axial potential distribution at $z=0$ or at any other point.

A comparative plot using various funcions is given in Fig. 20.

The relation between the object and image coordinates is obtained in the following manner: An electron path intersecting the optical axis at the object coordinate φ_0 and whose slope at this point is:

$$r_1'(z_0)=1,$$

can be expressed as:

$$r_1=-\frac{a_1}{\omega_1 \sin\varphi_0}\exp\left[\tfrac{1}{3}\sqrt{3}k_1(\varphi-\varphi_0)\right]\cdot\frac{\sin\left[\omega_1(\varphi-\varphi_0)\right]}{\sin\varphi}.$$

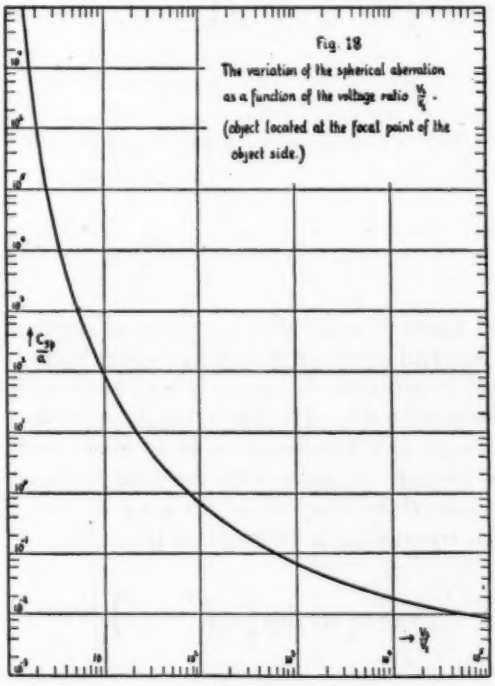
Fig. 18
The variation of the spherical aberration as a function of the voltage ratio $\frac{V_2}{V_1}$.
(object located at the focal point of the object side.)

FIG. 18.

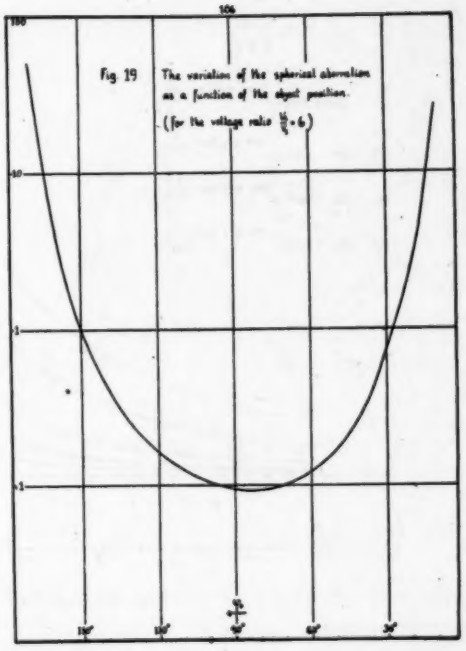

FIG. 19. The variation of the spherical aberration as a function of the object position (for the voltage ratio $V_2/V_1=6$).

The ordinate of this path at $z=0$ is then:

$$r_1|_{z=0}=-\frac{a_1}{\omega_1 \sin\varphi_0}\exp\left[\tfrac{1}{3}\sqrt{3}k_1\left(\frac{\pi}{2}-\varphi_0\right)\right]\cdot\frac{\sin\left[\omega_1\left(\frac{\pi}{2}-\varphi_0\right)\right]}{\sin\frac{\pi}{2}},$$

and its slope at the same point is:

$$r_1'|_{z=0}=\frac{1}{\omega_1 \sin\varphi_0}\exp\left[\tfrac{1}{3}\sqrt{3}k_1\left(\frac{\pi}{2}-\varphi_0\right)\right]\cdot\left[\tfrac{1}{3}\sqrt{3}k_1\sin\omega_1\left(\frac{\pi}{2}-\varphi_0\right)+\omega_1\cos\omega_1\left(\frac{\pi}{2}-\varphi_0\right)\right].$$

An electron path intersecting the optical axis at the image coordinate φ_i is given by:

$$r_2=C\exp\left[\tfrac{1}{3}\sqrt{3}k_2\left(\varphi-\frac{\pi}{2}\right)\right]\frac{\sin\left[\omega_2(\varphi-\varphi_i)\right]}{\sin\varphi}.$$

The constant can now be determined so that both paths have the same ordinate and the same slope at $z=0$, if φ_0 and φ_i are really object and

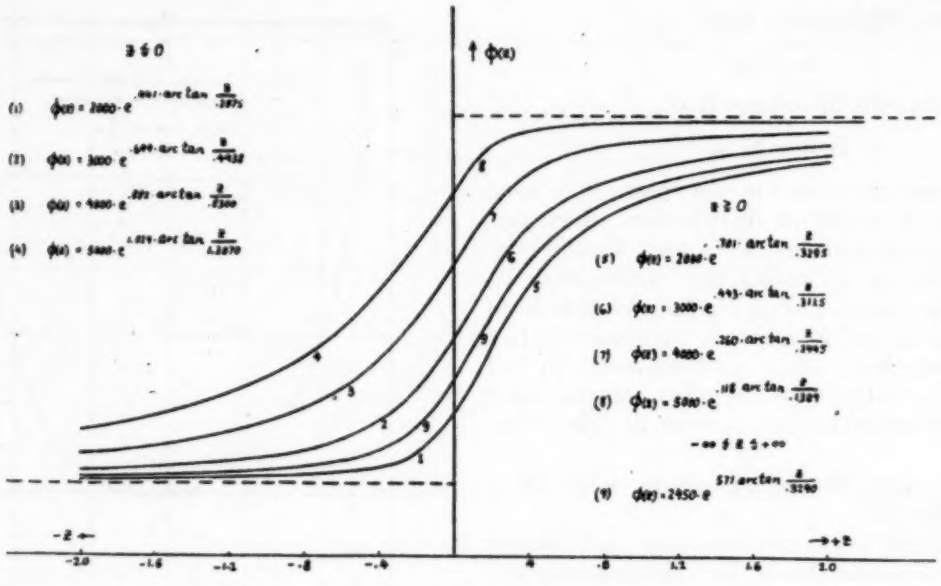

FIG. 20. Axial potentials of the generalized lens.

image coordinates. The matching conditions are:

$$C\sin\left[\omega_2\left(\frac{\pi}{2}-\varphi_i\right)\right]=-\frac{a_1}{\omega_1\sin\varphi_0}\times\exp\left[\tfrac{1}{3}\sqrt{3}k_1\left(\frac{\pi}{2}-\varphi_0\right)\right]\cdot\sin\left[\omega_1\left(\frac{\pi}{2}-\varphi_0\right)\right].$$

$$C\left\{\frac{\sqrt{3}}{3}k_2\sin\left[\omega_2\left(\frac{\pi}{2}-\varphi_i\right)\right]+\omega_2\cos\left[\omega_2\left(\frac{\pi}{2}-\varphi_i\right)\right]\right\}=\frac{a_2\exp\left[\frac{\sqrt{3}}{3}k_1\left(\frac{\pi}{2}-\varphi_0\right)\right]}{\omega_1\sin\varphi_0}\left\{\frac{\sqrt{3}}{3}k_1\sin\left[\omega_1\left(\frac{\pi}{2}-\varphi_0\right)\right]+\omega_1\cos\left[\omega_1\left(\frac{\pi}{2}-\varphi_0\right)\right]\right\}.$$

Division of the second equation by the first gives:

$$\tfrac{1}{3}\sqrt{3}k_2+\omega_2\cot\left[\omega_2\left(\frac{\pi}{2}-\varphi_i\right)\right]=\frac{a_2}{a_1}\left\{\tfrac{1}{3}\sqrt{3}k_1+\omega_1\cot\left[\omega_1\left(\frac{\pi}{2}-\varphi_0\right)\right]\right\},$$

or since:

$$k_2=\frac{a_2}{a_1}k_1,$$

$$\tan\left[\omega_2\left(\frac{\pi}{2}-\varphi_i\right)\right]=\frac{a_1\omega_2}{a_2\omega_1}\tan\left[\omega_1\left(\frac{\pi}{2}-\varphi_0\right)\right].$$

This is the relation between the object and image coordinates. For the case $a_1=a_2$ and $k_1=k_2$, it becomes identical with Eq. (26) as required.

Using Eq. (38) the matching constant C can be determined; it is:

$$C=-\frac{\exp\left[\tfrac{1}{3}\sqrt{3}k_1\left(\frac{\pi}{2}-\varphi_0\right)\right]}{\omega_1\omega_2\sin\varphi_0}\cdot\left\{a_1^2\omega_2^2\sin^2\left[\omega_1\left(\frac{\pi}{2}-\varphi_0\right)\right]+a_2^2\omega_1^2\cos^2\left[\omega_1\left(\frac{\pi}{2}-\varphi_0\right)\right]\right\}^{\frac{1}{2}}.$$

W. Glaser[5] has given a graphical method for the determination of the image coordinate if the object coordinate is given. This construction is illustrated in Fig. 21. The triangle AOC has the base $a_1\omega_2$, and the angle at A is made equal to $\omega_1(\pi/2-\varphi_0)$. If, then, OB is equal to $a_2\omega_1$, the angle at B will be equal to $\omega_2(\pi/2-\varphi_i)$. The image coordinate is then given by:

$$z_i=a_2\operatorname{arc\,cot}\left[\omega_2\left(\frac{\pi}{2}-\varphi_i\right)\right].$$

The magnification can be determined by the

same method which we used previously for the single function lens. A solution of the paraxial ray equation for the left-hand field which is different from zero at the object position and which is zero at $z=0$ is given by:

$$r_1=\exp\left[\tfrac{1}{3}\sqrt{3}k_1(\varphi-\pi/2)\right]\cdot\frac{\sin\left[\omega_1(\varphi-\pi/2)\right]}{\sin\varphi}.$$

The path function in the second part of the field is a continuous extension of that in the first part if the constants in:

$$r_2=C_1\exp\left[\tfrac{1}{3}\sqrt{3}k_2(\varphi-\pi/2)\right]\cdot\frac{\sin\left[\omega_2(\varphi-C_2)\right]}{\sin\varphi}$$

are:

$$C_1=a_2\omega_1/a_1\omega_2,\quad C_2=\pi/2.$$

The magnification is then nothing else but the following ratio:

$$M=r_1(z_0)/r_2(z_i).$$

Substitution of the particular solutions r_1 and r_2 yields:

$$M=\frac{k_2}{k_1}\frac{\omega_1}{\omega_2}\exp\left[\tfrac{1}{3}\sqrt{3}\{k_2(\varphi_i-\pi/2)-k_1(\varphi_0-\pi/2)\}\right]\cdot\frac{\sin\left[\omega_2(\varphi_i-\pi/2)\right]}{\sin\left[\omega_1(\varphi_0-\pi/2)\right]}\cdot\frac{\sin\varphi_0}{\sin\varphi_i}.$$

The two focal points can be determined by means of Eq. (38). If $\varphi_i=0$, i.e., $z=+\infty$ and $\varphi_0=\pi$, i.e., $z=-\infty$ they are given by the two following equations:

$$\left.\begin{aligned}\tan\left[\omega_2\left(\frac{\pi}{2}-\varphi_{F_i}\right)\right]&=-\frac{k_1}{k_2}\frac{\omega_2}{\omega_1}\tan\left[\omega_1\frac{\pi}{2}\right]\\ \tan\left[\omega_1\left(\frac{\pi}{2}-\varphi_{F_0}\right)\right]&=-\frac{k_2}{k_1}\frac{\omega_1}{\omega_2}\tan\left[\omega_2\frac{\pi}{2}\right]\end{aligned}\right\}.$$

We should like to emphasize that the notion of focal length loses its meaning for the lens described by the two expressions:

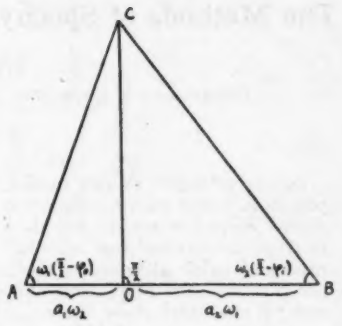

FIG. 21. Construction of the image coordinate.

$$\phi_1(z)=\phi_0\exp\left(\frac{4}{3}\sqrt{3}k_1\arctan\frac{z}{a_1}\right),$$

$$\phi_2(z)=\phi_0\exp\left(\frac{4}{3}\sqrt{3}k_2\arctan\frac{z}{a_2}\right)$$

since the latter does not belong to the class of lenses which satisfy Newton's image relations. In this case the simple image construction cannot therefore be used.

Spherical and chromatic aberration can also be computed for the lens given by Eqs. (38a). This can be done by integrating the expressions for C_{chr} and C_{sp} in two steps. The first integral has the limits φ_0 and $\pi/2$, the second integral is taken between $\pi/2$ and φ_i. The influence of different ratios k_2/k_1 on the magnitude of the aberrations could be investigated. This investigation, however, would be beyond the scope of this work and shall therefore be omitted.

ACKNOWLEDGMENT

The present work was carried out while the author was connected with the Division of Electron Optics at Stanford University. For valuable advice the author is indebted to Professor L. Marton under whose direction the work was done. Discussions with Professors G. Szego, G. Polya, and W. W. Hansen of Stanford University were helpful in obtaining solutions of a number of mathematical problems. The numerical calculations were carried out by my wife, Ruth S. Hutter.

The Methods of Specirying the Properties of Viscoelastic Materials

TURNER ALFREY AND PAUL DOTY
Department of Chemistry, Polytechnic Institute of Brooklyn, Brooklyn, New York
(Received May 17, 1945)

Seven methods (Voigt model, Maxwell model, operator equation, mechanical impedance function, creep curve, relaxation curve, and dynamic modulus function) of specifying viscoelastic behavior are discussed. A number of exact relations between these methods of specification are worked out in detail. The majority of these relations are simple enough to be of practical value although a few are too cumbersome. Approximate relationships between the creep curve, the relaxation curve, Maxwell model, and Voigt model are discussed; and numerical examples show the magnitude of errors introduced by the approximation to be small even in quite unfavorable cases. A consideration of the practical utility and physical meaning of the various methods of specification distinguishes between (1) those of general descriptive value and those of direct experimental value; (2) those useful in a phenomenological study of mechanical behavior and those more suited to a formulation of molecular theory. A summary of the present molecular theories is presented together with their interpretation in terms of the Voigt and Maxwell specifications.

INTRODUCTION

DURING the last decade the investigation of the mechanical behavior of viscoelastic materials has greatly expanded. Accompanying this development there has been a corresponding increase in the number of methods used by various investigators to summarize their data and to specify the properties of the materials studied. In many cases it has not been clearly recognized that the methods used were actually equivalent and this has led to an incomplete utilization of the data available. No less than seven distinct methods have been employed to define the properties (in shear) of viscoelastic materials. It is the purpose of this paper to examine these methods and show that they are all mathematically equivalent and that they are all interconnected by definite relationships.

We are only concerned with *incompressible, isotropic, viscoelastic* materials. The term incompressible indicates that the material is much more easily deformed in shear than in compression; thus, the viscoelastic deformation is essentially a constant volume process. The term isotropic requires that in the unstrained state the properties of the material are the same in all directions. The term viscoelastic is here restricted to designate those materials whose response to stress obeys the superposition principle;[1] that is, materials which can be represented by a model constructed from elements which obey Hooke's elastic law and elements which obey Newton's viscosity law. The mechanical properties of such an incompressible, isotropic, viscoelastic material are completely defined by its response to homogeneous shearing stresses. A rigorous method is available for calculating the behavior in the presence of combined and inhomogeneous stresses.[2] The mathematical theory of non-linear strain has been developed by Murnaghan.[3]

The seven distinct methods of specifying the properties of viscoelastic materials may be divided into two classes. In Class I may be listed the four that are more general and fundamental in nature.

A. The Voigt model, consisting of a set of Voigt elements (retarded elastic elements) coupled in series.
B. The Maxwell model, consisting of a set of Maxwell elements coupled in parallel.
C. The operator equation, $Ps=2Q\epsilon$, where **P** and **Q** are linear differential operators.
D. The mechanical impedance function, $Z(\omega)$.

In Class II may be listed the obvious experimental curves which serve to "map out" the viscoelastic character of the material.

E. The creep curve, consisting of the strain as a function of time at constant stress.
F. The relaxation curve, consisting of the stress as a function of time at constant strain.
G. The dynamic modulus curve, consisting of the elastic modulus as a function of frequency.

The Class I methods are more satisfactory as specifications of properties since they are of the nature of differential equations which can be solved for a wide variety of transient conditions. Class II methods on the other hand represent special solutions of the fundamental differential

[1] J. C. Maxwell, Phil. Mag. **35**, 134 (1868).

[2] T. Alfrey, Quar. App. Math. **2**, 113 (1944).
[3] F. D. Murnaghan, Am. J. Math. **59**, 235 (1937).

equations for very particular cases. Consequently, these latter functions are not so well adapted to the treatment of a wide range of transient conditions.

It is obviously important to be able to transform readily from one fundamental mode of description to another, to be able to derive the basic properties of a material (Class I) from experimental data (Class II), to be able to predict behavior in special conditions (Class II) from the properties (Class I), and to be able to predict behavior in one special set of conditions from the known behavior under other conditions. If Class II methods are accepted as specifying the properties of a viscoelastic material, then all of these problems are concerned with the transformation from one of the seven modes of description to another.

Some of these transformations have been worked out and are of a form simple enough to be of practical value. In Fig. 3 these transformations are indicated and labeled as a, b, $\cdots$, m. Several other transformations have been formally developed but are perhaps too cumbersome to serve as practical working schemes. These are indicated in Fig. 4 and labeled n, o, $\cdots$, t. Finally, several transformations can be made by simple approximation methods. These are also summarized in Fig. 4 and labeled u, v, w, and x.

This paper is divided into four parts:

I. Brief, individual discussions of the seven methods of specifying viscoelastic properties.
II. Development of exact relationships among the seven methods.
III. Development of approximate relationships among the methods.
IV. Discussion of the relative utility and physical meaning of the different methods.

I. METHODS OF SPECIFYING VISCOELASTIC PROPERTIES

A. The Voigt Model

A Voigt element[4] consists of a spring (representing Hooke's law behavior) and a dashpot (representing Newton's viscosity law behavior) in parallel. This element expresses the simplest form of retarded elastic behavior. Such a single Voigt element is characterized by two parameters—the modulus of the spring, G, and the viscosity of the dashpot, η. The strain, 2ϵ, associated with the element follows the differential equation.

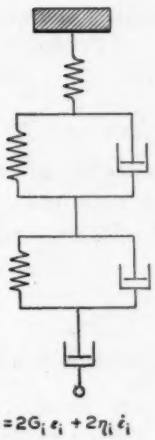

FIG. 1. The Voigt model.

$$s=2G\epsilon+2\eta\dot{\epsilon}, \tag{1}$$

where s is the stress. The factor 2 in this and following equations arises from the fact that our ϵ represents a shearing component of the strain tensor, and is thus equal to $\frac{1}{2}$ of the corresponding "classical" shearing strain. It follows that s represents the corresponding component of the deviatoric stress tensor.* In the case of simple shear, s is the shearing stress and 2ϵ is the shearing strain.

An important quantity is τ, the retardation time of the element, which is equal to η/G. Another useful quantity is the compliance, J, which is simply the reciprocal of the modulus G. It will be convenient to use G and J interchangeably as a measure of the equilibrium elastic properties.

Figure 1 shows a Voigt model consisting of four Voigt elements, one of which is degenerated into a spring with no dashpot and one into a dashpot with no spring. The material is thus defined by the values of six parameters, J_1, J_2, η_2, J_3, η_3, and η_4. The total strain, ϵ, is made up of the contributions from the four elements: $\epsilon=\epsilon_1+\epsilon_2+\epsilon_3+\epsilon_4$. The retarded elastic elements obey the differential equations:

$$s=2G_2\epsilon_2+2\eta_2\dot{\epsilon}_2, \tag{2a}$$

$$s=2G_3\epsilon_3+2\eta_3\dot{\epsilon}_3. \tag{2b}$$

[4] W. Voigt, *Lehrbuch der Krystallphysik* (Leipzig, 1928).

* The stress tensor σ_{ij}, can always be resolved into two parts—the mean normal stress σ, and the deviatoric stress tensor s_{ij}. The mean normal stress is equal to $\frac{1}{3}(\sigma_{11}+\sigma_{22}+\sigma_{33})$ and s_{ij} is given by $\sigma_{ij}-\sigma\delta_{ij}$, where δ is the Kroneker delta. In the case of all isotropic materials this is in fact a very useful resolution since it permits the isolation of pressure-volume effects and shear effects. The deviatoric stress tensor includes all shearing stresses and only these. In an incompressible isotropic material the role of the deviatoric stress tensor is even more important since no strain results from the mean normal stress.

For the degenerate elements these of course reduce to

$$s=2G_1\epsilon_1, \tag{3a}$$

$$s=2\eta_4\dot{\epsilon}_4. \tag{3b}$$

More generally, the behavior of a material which is represented by n retarded elastic elements is governed by a set of n-differential equations of the form

$$s=2G_i\epsilon_i+2\eta_i\dot{\epsilon}_i, \tag{4}$$

(of which one or two may be degenerate) plus the relation $\epsilon=\sum^{n}\epsilon_i$. In the limit, the set of discrete Voigt elements may be replaced by a continuous distribution of elastic compliance as a function of retardation time. That is, the finite set of physical constants—J_1, η_1, J_2, η_2, $\cdots$, J_n, η_n—is now replaced by a continuous function $J(\tau)d\tau$ which tells how much elastic compliance has the retardation time, τ, in the range of $d\tau$. The total retarded strain is given by the integral,

$$\epsilon(t)=\int_0^\infty \epsilon^*(t,\tau)d\tau,$$

where ϵ^* is governed by the differential equation:

$$s(t)=\frac{2\epsilon^*(t,\tau)}{J(\tau)}+\frac{2\tau}{J(\tau)}\frac{\partial\epsilon^*(t,\tau)}{\partial t}. \tag{5}$$

B. The Maxwell Model

Figure 2 shows a Maxwell model[1] consisting of four Maxwell elements, one of which is degenerated to a spring and one to a dashpot. In this model the strains associated with all elements are equal, but the total stress is divided among the elements.[5,6] The stress on the ith element is related to the strain by the differential equation:

$$2\dot{\epsilon}=\frac{1}{G_i}\dot{s}+\frac{1}{\eta_i}s_i \tag{6}$$

which takes on simpler form for the two degenerated elements. As in the previous method, the behavior of the material is governed by a set of first-order differential equations.

In the limit, this behavior goes over into a continuous distribution[7] of elastic moduli as a function of relaxation time. The finite set of physical constants G_1, $\tau_1\cdots G_n$, τ_n is replaced by a function $G(\tau)d\tau$, telling the amount of elastic modulus with the relaxation time τ in the range $d\tau$. The total stress is given by the integral:

$$s(t)=\int_0^\infty s^*(t,\tau)d\tau,$$

where $s^*(t,\tau)$ follows the differential equation

$$2\dot{\epsilon}(t)=\frac{1}{G(\tau)}\frac{\partial s^*(t,\tau)}{\partial t}+\frac{s^*(t,\tau)}{\tau G(\tau)}. \tag{7}$$

C. The Operator Equation Ps—$2Q\epsilon$

Method A involved a set of n differential equations of the first order, each relating a particular strain contribution with the total stress. Method B involved a set of n differential equations of the first order, each relating a fraction of the stress with the total strain. Method C[2,8] involves a single differential equation of the nth order, relating the total stress with the total strain:

$$\mathbf{P}s=2\mathbf{Q}\epsilon \tag{8}$$

where

$$\mathbf{P}=\frac{\partial^m}{\partial t^m}+p_{m-1}\frac{\partial^{m-1}}{\partial t^{m-1}}+\cdots+p_0,$$

$$\mathbf{Q}=q_n\frac{\partial^n}{\partial t^n}+q_{n-1}\frac{\partial^{n-1}}{\partial t^{n-1}}+\cdots+q_0.$$

The parameters p_i, q_i, fix the properties of the viscoelastic material. It will be demonstrated in Part II that the same number of independent parameters are required to characterize a material, regardless of which method (A, B, or C) is used.

In a material which is characterized by a continuous distribution of elastic compliance $J(\tau)$

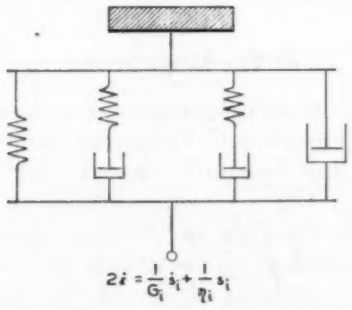

FIG. 2. The Maxwell model.

[5] K. Bennewitz and H. Rotger, Physik. Zeits. **40**, 416 (1939).
[6] W. Holzmuller and E. Jenckel, Zeits. f. physik. Chemie **A186**, 359 (1940).
[7] W. Kuhn, Zeits. f. physik. Chemie **B42**, 1 (1939).
[8] T. Alfrey, *Mechanical Properties of High Polymers* (Interscience Publishers, Inc., New York, 1946).

instead of a model with a finite number of discrete elements, the operators **P** and **Q** likewise must take the forms of integrals rather than sums.

$$\mathbf{P}=\int_0^\infty p(x)\frac{\partial^x}{\partial t^x}dx,$$

$$\mathbf{Q}=\int_0^\infty q(x)\frac{\partial^x}{\partial t^x}dx.$$

The properties are now fixed by the continuous functions $p(x)$ and $q(x)$ instead of a set of constants p_i, q_i.

D. The Mechanical Impedance Function

The analogy between electrical networks and the mechanical models designed to describe viscoelastic behavior seems quite obvious once it is recognized. Voltage corresponds to stress, resistance to viscosity, current to the time derivative of strain and the reactance to $-G/\omega$. Consequently, the steady-state response of a viscoelastic material to a sinusoidal stress $s=s_0\cos\omega t$ can be related to a (complex) mechanical impedance Z. $Z=a+ib$, where a is the "viscous resistance," η, and b the "elastic reactance," $-G/\omega$. Both a and b depend in magnitude upon the frequency. The complex mechanical impedance function $Z(\omega)$ fixes the properties of a viscoelastic material in a manner completely equivalent to those of A, B, and C. Although the most obvious use of the impedance function is in connection with steady-state response to sinusoidal stress, the Fourier and Laplace transformations make it as general as A, B, and C in its definition of viscoelastic properties.

E. The Creep Function

If a relaxed specimen of a viscoelastic material is suddenly subjected (at the time $t=0$) to a stress s which thereafter is held constant, the strain will increase with time according to some function $2\epsilon(t)$. Since the strain at any time is proportional to the stress s, dividing through by s gives a creep function $\phi(t)$ which is independent of the stress. That is,

$$\phi(t)=2\epsilon(t)/s.$$

A given creep function is consistent with only one model of either type and with only one operator specification. Although the creep function can be said to define the viscoelastic properties of a material, as has been pointed out, it is less satisfactory than A, B, C, and D for such a specification.

F. The Relaxation Function

If a sample is quickly forced to take on the strain ϵ, and is then held fixed in such a fashion that the strain must remain constant, the stress will decay with time according to some function $s(t)$.[9,10] Since the stress at any time is proportional to the strain, dividing through by the strain gives a relaxation function $\psi(t)$ which is independent of the strain. That is

$$\psi(t)=s(t)/2\epsilon.$$

The relaxation function of a material defines its viscoelastic properties in precisely the same sense as does the creep function.

G. The Dynamic Modulus

When subjected to a stress which varies sinusoidally with time ($s=s_0\cos\omega t$), a viscoelastic material will respond[11] with a strain $2\epsilon(t)$ which (in the steady state) will also vary sinusoidally with time: $2\epsilon=2\epsilon_0\cos(\omega t-\alpha)$. The ratio, $s_0/2\epsilon_0$, can be called the dynamic modulus, $G(\nu)$, and will in general be a function of the frequency. A given dynamic modulus function will be consistent with only one model of either type, and with only one operator specification. Hence, $G(\nu)$ can be said to specify the viscoelastic properties of a material.

II. EXACT RELATIONSHIPS AMONG THE METHODS

While any one of the above methods is in principle sufficient to define the properties of a given viscoelastic material, it is desirable to understand how the various methods are related to one another. This is particularly true at the moment. The wide differences in choice of method to be found in the current literature often make it difficult to see the precise connection between the results of two different investigators in the same field. This lack of agreement on the language to be used is a healthy, if confusing, index of the present status of the theory of viscoelastic behavior; it makes imperative, however,

[9] J. J. Thomson, *Applications of Dynamics to Physics and Chemistry* (London, 1888).
[10] E. Wiechert, Ann. Chim. Phys. **50**, 335 (1893).
[11] A. P. Alexandrov and Yu. S. Lazurkin, Acta Physicochimica **12**, 647 (1940).

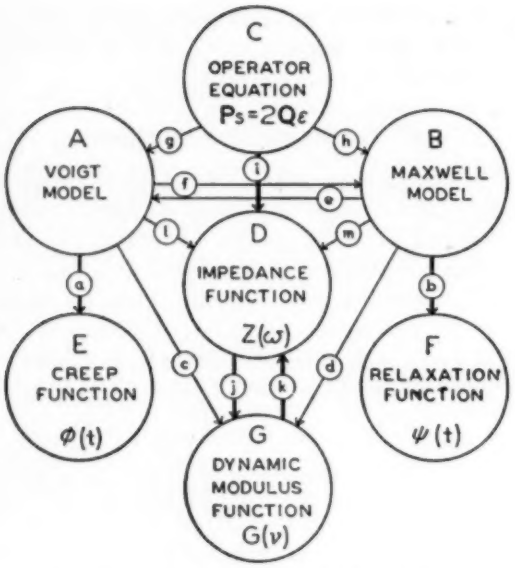

FIG. 3. Diagram of practical relationships.

a ready means of translation from one actively used language to another.

In this section, a number of exact relationships will be discussed. These may be divided into two classes—those which are simple enough to be used as working methods, and those which are chiefly formal in nature. The diagram in Fig. 3 shows the translations which can be made exactly and yet fairly simply; in Fig. 4 the formal relationships, more difficult to utilize, are indicated together with approximate relationships. The lettering in these figures will correspond to that in this section.

A. Voigt Model to Creep Function

If the viscoelastic properties of a material have been specified by a Voigt model, it is always easy to compute the response of the material, $2\epsilon(t)$, to any simple stress sequence $s(t)$, since the response of each Voigt element can be calculated separately, and then the individual contributions to the strain can be added together. The creep curve is a particularly simple case. A constant stress s, imposed upon a single relaxed Voigt element at time zero, causes a strain response, $2\epsilon(t)$, which must obey the differential equation

$$2G\epsilon+2\eta\dot{\epsilon}=s, \tag{9}$$

where $s=$constant. The general solution of this equation is

$$2\epsilon(t)=Ae^{-t/\tau}+(s/G). \tag{10}$$

Inserting the initial condition that $\epsilon=0$ at time $t=0$ for a sample containing no residual strains we can evaluate the arbitrary constant, A, and obtain the particular solution

$$2\epsilon(t)=\frac{s}{G}[1-e^{-t/\tau}]=sJ[1-e^{-t/\tau}]. \tag{11}$$

If the model consists of n retarded elastic elements, each with its own compliance and retardation time, the total response of the model is given by

$$2\epsilon(t)=s\sum_{i=1}^{n}J_i[1-e^{-t/\tau_i}]. \tag{12}$$

If in addition to n complete Voigt elements the model includes an unretarded compliance J_1 and a true viscous flow governed by the viscosity η, the response becomes

$$2\epsilon(t)=s\left[J_1+\sum_{i=2}^{n}J_i(1-e^{-t/\tau_i})+\frac{1}{\eta}t\right]. \tag{13}$$

Finally, if the series n Voigt elements is replaced by a continuous distribution of elastic compliance, $J(\tau)$, over a range of τ, Eq. (13) becomes

$$2\epsilon(t)=s\left[J_1+\int_0^\infty J(\tau)(1-e^{-t/\tau})d\tau+\frac{1}{\eta}t\right]. \tag{14}$$

Each of these functions can be changed to the standard creep function $\phi(t)$ simply by cancelling out the constant s from the right-hand side.

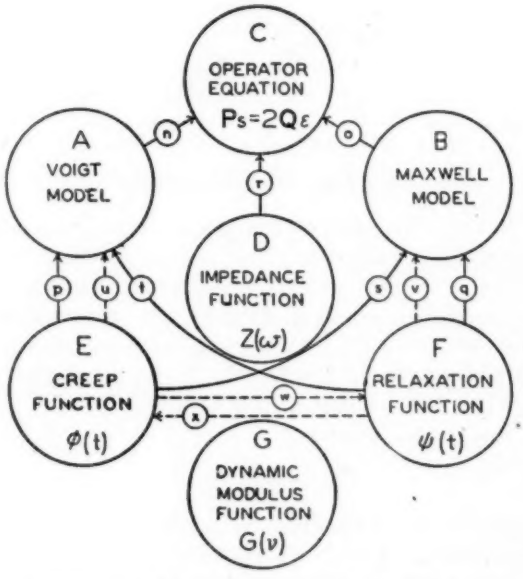

FIG. 4. Diagram of formal (——) and approximate (---) relationships.

B. Maxwell Model to Relaxation Function

While the Voigt specification allows a simple prediction of the strain as a function of time for a known $s(t)$, the Maxwell model is designed to allow an equally simple prediction of the stress as a function of time when a known strain sequence, $2\epsilon(t)$, is forced upon the material. $s(t)$ for the known $2\epsilon(t)$ can be computed for each Maxwell element separately and the results added together. The relaxation of stress at constant strain is a particularly simple case. (It is here assumed that the strain 2ϵ is suddenly imposed upon a relaxed sample at $t=0$.) The stress in a given Maxwell element must follow the differential equation:

$$2\dot{\epsilon}=(1/G)\dot{s}+(1/\eta)s. \tag{15}$$

This leads to the explicit relation:

$$s(t)=s_0e^{-(G/\eta)t}=2G\epsilon e^{-t/\tau}. \tag{16}$$

For a set of n Maxwell elements in parallel and a spring of modulus G, with no dashpot for relaxation, Eq. (16) becomes

$$s(t)=2\epsilon\left[G_1+\sum_{i=2}^{n}G_ie^{-t/\tau_i}\right]. \tag{17}$$

We need not consider here the case where an isolated dashpot is present. Finally, if the set of n Maxwell elements is replaced by a continuous distribution of relaxation times, $G(\tau)$, the stress relaxation at constant strain follows the relation

$$s(t)=2\epsilon\int_0^\infty G(\tau)e^{-t/\tau}d\tau. \tag{18}$$

C. Voigt Model to Dynamic Modulus Function

The steady-state relationship between sinusoidal stress and strain is unique in one respect. Since this relationship can be reduced to a vectorial relation, between the two members of a rotating rigid pair of vectors, either $s(t)$ or $2\epsilon(t)$ can be taken as the "known" or "impressed" function, and the relative amplitude and phase of the other computed. If the viscoelastic properties are specified by a Voigt model, we obviously must choose $s(t)$ as the "known" function. $S(t)=S_0(\cos\omega t)$. A single Voigt element must respond to this stress according to the differential equation

$$s_0\cos\omega t=2G\epsilon+2\eta\dot{\epsilon}. \tag{19}$$

The steady-state solution is

$$2\epsilon=s_0\frac{\sin\left(\omega t-\tan^{-1}\dfrac{-1}{\omega\tau}\right)}{G(\omega^2t^2+1)^{\frac{1}{2}}}. \tag{20}$$

A set of n Voigt elements, plus an unretarded compliance J_1 and a viscous flow term η, will therefore have a steady-state response equal to

$$2\epsilon=s_0\left[J_1\cos\omega t+\sum\frac{\sin\left(\omega t-\tan^{-1}\dfrac{-1}{\omega\tau_i}\right)}{G_i(\omega^2\tau_i^2+1)^{\frac{1}{2}}}+\frac{1}{\eta\omega}\sin\omega t\right]. \tag{21}$$

The extension to a continuous distribution of relaxation times is obvious.

D. Maxwell Model to Dynamic Modulus Function

When a Maxwell specification is used, we must choose the reverse formulation of the problem, and let $2\epsilon=2\epsilon_0\sin\omega t$. By obvious steps we obtain, in the case of a set of n Maxwell elements,

$$s=2\epsilon_0\left[\sum^{n}\frac{\omega^2\tau_i^2G_i}{1+\omega^2\tau_i^2}\sin\omega t+\sum^{n}\frac{\omega\tau G_i}{1+\omega^2\tau^2}\cos\omega t\right], \tag{22}$$

and in the case of a continuous distribution of elastic relaxation times.

$$s=A\sin\omega t+B\cos\omega t, \tag{23}$$

where

$$A=2\epsilon_0\omega^2\int_0^\infty\frac{\tau^2G(\tau)}{1+\omega^2\tau^2}d\tau,\quad B=2\epsilon_0\omega\int_0^\infty\frac{\tau G(\tau)}{1+\omega^2\tau^2}d\tau.$$

The dynamic modulus function is then given by $G(\omega)=(A^2+B^2)^{\frac{1}{2}}$.

E, F. Maxwell Model to and from Voigt Model

The translations from a Voigt model to a Maxwell model, and *vice versa*, have been indicated in Fig. 3 as being relatively simple. This is only partly true. If we are given a Voigt model, and must construct the equivalent Maxwell model, the first step is to find the correct struc-

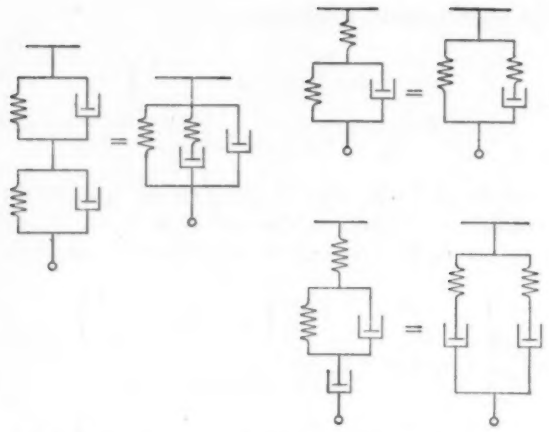

FIG. 5. Equivalence of Voigt and Maxwell models.

tural *form* of the unknown model, and the second is to compute the numerical values of the unknown parameters. The first, or topological, step is controlled by rules which can be formulated simply, but the second step can be easily made only in very simple cases.

The following rules govern the topological relationships between the two kinds of model:

1. The Maxwell model will contain the same number of springs and the same number of dashpots as the equivalent Voigt model.

2. A standard Voigt model (no incomplete elements) corresponds to a "doubly degenerate" Maxwell model (one isolated spring and one isolated dashpot).

3. A standard Maxwell model corresponds to a "doubly degenerate" Voigt model.

4. A singly degenerate Voigt model with one isolated (unretarded) spring corresponds to a singly degenerate Maxwell model with one isolated (non-relaxing) spring.

5. A singly degenerate Voigt model with one isolated dashpot corresponds to a singly degenerate Maxwell model with one isolated dashpot.

6. The rather involved general rule which embraces 2, 3, 4, and 5 is the following: The *presence* of an isolated element of *one* kind in a model corresponds to the *absence* of an isolated element of the *opposite* kind in the conjugate model (e.g., the presence of an isolated *spring* in a Voigt model implies the *absence* of an isolated *dashpot* in the equivalent Maxwell model).

Figure 5 shows a number of specific illustrations of these generalizations.

The numerical computation of the parameters of one model from those of its equivalent is much more involved. It is, indeed, out of the question to write down general functional relationships for one set in terms of the other. The complexity which arises can be estimated from the following explicit equations for the relatively simple case of two Maxwell elements in parallel, corresponding to one unretarded spring, one Voigt element, and one dashpot in the other language. Let G_1^M, η_1^M, G_2^M, and η_2^M be the parameters of the Maxwell model, and G_1^V, G_2^V, η_2^V, η_3^V the parameter of the equivalent Voigt model. The transformation equations are:

$$G_1^V = G_1^M + G_2^M, \tag{24a}$$

$$G_2^V = \frac{G_1^M G_2^M (\eta_1^M + \eta_2^M)^2 (G_1^M + G_2^M)}{(\eta_1^M G_2^M - \eta_2^M G_1^M)}, \tag{24b}$$

$$\eta_2^V = \frac{\eta_1^M \eta_2^M (G_1^M + G_2^M)^2 (\eta_1^M + \eta_2^M)}{(\eta_1^M G_2^M - \eta_2^M G_1^M)}, \tag{24c}$$

$$\eta_3^V = \eta_1^M + \eta_2^M, \tag{24d}$$

$$G_1^M = \frac{1}{2\beta}\left(G_1^V\alpha + G_1^V\beta - \frac{(G_1^V)^2}{\eta_2^V} - \frac{(G_1^V)^2}{\eta_3^V}\right), \tag{25a}$$

$$G_2^M = \frac{-1}{2\beta}\left(G_1^V\alpha - G_1^V\beta - \frac{(G_1^V)^2}{\eta_2^V} - \frac{(G_1^V)^2}{\eta_3^V}\right), \tag{25b}$$

$$\eta_1^M = \frac{1}{2\beta}\left(G_1^V\alpha + G_1^V\beta - \frac{(G_1^V)^2}{\eta_2^V} - \frac{(G_1^V)^2}{\eta_3^V}\right)\left(\frac{1}{\alpha+\beta}\right), \tag{25c}$$

$$\eta_2^M = \frac{-1}{2\beta}\left(G_1^V\alpha - G_1^V\beta - \frac{(G_1^V)^2}{\eta_2^V} - \frac{(G_1^V)^2}{\eta_3^V}\right)\left(\frac{1}{\alpha+\beta}\right), \tag{25d}$$

where

$$\alpha = \left(\frac{G_1^V}{2\eta_3^V} + \frac{G_1^V}{2\eta_2^V} + \frac{G_1^V}{2\eta_2^V}\right), \tag{26a}$$

$$\beta = \alpha^2 - \frac{G_2^V G_1^V}{\eta_2^V \eta_3^V}. \tag{26b}$$

Because of this complexity, it is suggested that numerical computations of this sort be made in two steps, via the operator formulation, rather than directly.

G. Operator Equation to Voigt Model

The first step in translating an operator equation $\mathbf{P}s = 2\mathbf{Q}\epsilon$ into the equivalent Voigt model is to establish the correct structural form (topology) of the model. The following rules govern the form of the model.

1. The total number of ultimate elements of the model is equal to the total number of parameters involved in the operators **P** and **Q**.

2. If the Voigt model contains no isolated springs or dashpots, the order of **P** is one less than that of **Q**. ($n=m+1$.) The "standard" operator equation which is equivalent to a set of n complete Voigt elements can therefore be written:

$$\frac{\partial^{n-1}s}{\partial t^{n-1}}+p_{n-2}\frac{\partial^{n-2}s}{\partial t^{n-2}}+\cdots+p_0 s=2q_n\frac{\partial^n\epsilon}{\partial t^n}+\cdots+2q_0\epsilon. \quad (27)$$

3. If the coefficient q_n vanishes, then one Voigt element of the equivalent model degenerates to a spring with no dashpot.

4. If the coefficient q_0 vanishes, one Voigt element of the equivalent model degenerates to a dashpot with no spring.

5. If q_0 and q_n both vanish, the corresponding Voigt model is *doubly* degenerate.

The second step of the problem is the determination of numerical values for the parameters of the model. This step is discussed in detail in another paper.[12] The standard, or non-degenerate, case will be briefly outlined here.

It is clear that in order for the operator equation and the Voigt model to be mathematically equivalent, they must give the same prediction for the strain, $\epsilon(t)$, when any given stress sequence $s(t)$ is imposed. It is sufficiently general to equate the two responses to a stress sequence $s(t)=t^{n-1}$. The general solution of the equation $\mathbf{P}s=2\mathbf{Q}\epsilon$ is the sum of the general solution of the associated homogeneous equation $s(t)=0$ and the particular polynomial solution of the complete equation. In the same way, the response of the model to the stress $s(t)$ is the sum of the general response to a zero stress plus a particular polynomial response to the stress t^{n-1}. In order for the response of the model to be identical with that predicted by the operator equation, the parameters of the model must fit the following rules. First, the n retardation times of the n Voigt elements are the negative reciprocals of the n roots of the characteristic equation of the operator **Q**; that is $\tau_i=-1/x_i$, where x_i are the roots of the equation $q_n x^n+\cdots+q_0=0$. Thus, the n retardation times of the model are determined by the general solution of the homogeneous differential equation.

In order to complete the specification of the model the particular polynomial solution must now be used. The particular polynomial solution of the equation $\mathbf{P}(t^{n-1})=2\mathbf{Q}\epsilon$ will be of the form:

$$2\epsilon(t)=a_0+a_1t+a_2t^2+\cdots+a_{n-1}t^{n-1}. \quad (28)$$

The coefficients $a_0, a_1, \cdots a_{n-1}$ can be determined by ordinary methods. The particular polynomial solution derived from the model will be of the form:

$$2\epsilon(t)=(n-1)!\sum^n\frac{J_i}{x_i^{n-1}}+\left(\frac{(n-1)!}{2!}\sum^n\frac{J_i}{x_i^{n-2}}\right)t+\cdots$$
$$+\left(\frac{(n-1)!}{(n-2)!}\sum^n\frac{J_i}{x_i}\right)t^{n-2}+\sum^n J_i t^{n-1}. \quad (29)$$

It follows that the n compliances $J_1, J_2, \cdots J_n$ must be the roots of the following set of linear algebraic equations:

$$\begin{aligned}
a_{n-1}&=\sum^n J_i,\\
\frac{1}{n-1}a_{n-2}&=\sum\frac{J_i}{x_i},\\
\frac{1}{(n-1)(n-2)}a_{n-3}&=\sum\frac{J_i}{x_i^2},\\
&\cdots\cdots\cdots\\
\frac{1}{(n-1)!}a_0&=\sum\frac{J_i}{x_i^{n-1}}.
\end{aligned} \quad (30)$$

The Voigt model is completely specified when the compliance and retardation time of each Voigt element are determined.

The degenerate cases where q_n or q_0 vanish can also be simply handled.[12]

H. Operator Equation to Maxwell Model

A Maxwell model with no degenerate elements corresponds to an operator equation in which both q_n and q_0 vanish. The m relaxation times of the m Maxwell elements are equal to the negative reciprocals of the m roots of the characteristic equation of the operator **P**. The specification of the model can then be completed by the solution of a set of m linear equations in m unknowns.

$$\begin{aligned}
\frac{1}{m}a_{m-1}&=\sum^m\eta_i,\\
\frac{1}{m(m-1)}a_{m-2}&=\sum^m\frac{\eta_i}{x_i},\\
\frac{1}{m(m-1)(m-2)}a_{m-3}&=\sum^m\frac{\eta_i}{x_i^2},\\
&\cdots\cdots\cdots\\
\frac{1}{m!}a_0&=\sum^m\frac{\eta_i}{x_i^{m-1}}.
\end{aligned} \quad (31)$$

[12] T. Alfrey, Quar. App. Math. **3** (July, 1945).

I. Operator Equation to Impedance Function

In electrical network theory it is shown that the complex impedance, $Z(\omega)$, is equal to the ratio of the Fourier transform of the potential to the transform of the current. Since the current corresponds to the time derivative of the strain it is necessary to note the equivalent expression for the Fourier transform of $\dot{\epsilon}$; that is

$$\int_{-\infty}^{\infty}\dot{\epsilon}(t)e^{-i\omega t}dt=(i\omega)\int_{-\infty}^{\infty}\epsilon(t)e^{-i\omega t}dt. \tag{32}$$

$Z(\omega)$ can now be calculated from the operators **P** and **Q** as follows:

$$\mathbf{P}s=2\mathbf{Q}\epsilon, \tag{33}$$

$$\ddot{s}_2+p\dot{s}_1+p_0s=2[q_3\dddot{\epsilon}+q_2\ddot{\epsilon}+q_1\dot{\epsilon}+q_0\epsilon].$$

The Fourier transforms of each term are found by integrating over t after multiplying by $e^{i\omega t}$. Using (32), this procedure yields after collecting terms

$$[(i\omega)^2+p_1(i\omega)+p_0]\int_{-\infty}^{\infty}se^{-i\omega t}dt$$

$$=2[q_3(i\omega)^3+q_2(i\omega)^2+q_1(i\omega)+q_0]\int_{-\infty}^{\infty}\epsilon e^{-i\omega t}dt. \tag{34}$$

Thus the form of the impedance function is found to be:

$$Z(\omega)=\frac{\int s(t)e^{-i\omega t}dt}{(i\omega)2\int\epsilon(t)e^{-i\omega t}dt}$$

$$=\frac{q_3(i\omega)^3+q_2(i\omega)^2+q_1(i\omega)+q_0}{(i\omega)[(i\omega)^2+p_1(i\omega)+p_0]}$$

$$=\frac{\mathbf{Q}(i\omega)}{(i\omega)\mathbf{P}(i\omega)}. \tag{35}$$

J. Impedance Function to Dynamic Modulus

The relation between the impedance function and the response to sinusoidal stresses is, of course, particularly simple. At a given frequency, $\omega/2\pi$, where the mechanical admittance, $1/Z(\omega)$, has the value $A+iB$, the steady-state response to the stress, $s=s_0\cos\omega t$, will be:

$$2\dot{\epsilon}(t)=s_0A\cos\omega t-s_0B\sin\omega t, \tag{36a}$$

$$2\epsilon(t)=\frac{s_0}{\omega}A\sin\omega t+\frac{s_0}{\omega}B\cos\omega t. \tag{36b}$$

In this connection, the relationship between sinusoidal stress and strain (in the presence of a lumped mass) is sometimes written as:

$$M\ddot{\epsilon}+A\dot{\epsilon}+B\epsilon=s_0\cos\omega t, \tag{37}$$

or as some obvious variant of this equation. It is quite true that such an equation will hold for a particular frequency. In order for it to hold at *any* frequency, A and B must become functions of frequency, rather than numerical constants. This has, of course, been recognized by those using Eq. (37). It is, however, possible to write a differential equation relating stress and strain in which all coefficients are true constants. This is the operator equation, $\mathbf{P}s=2\mathbf{Q}\epsilon$, which for this problem (after the introduction of a lumped mass) becomes:

$$M\ddot{\epsilon}+2\mathbf{Q}\epsilon=\mathbf{P}(s_0\cos\omega t). \tag{38}$$

K. Dynamic Modulus to Impedance Function

The reverse transformation, by which the impedance function is derived from experimental data on the steady-state relation between sinusoidal stress and strain, is also simple. If the amplitude and phase of the strain relative to the stress, $s=s_0\cos\omega t$, is known, the strain can always be written as

$$2\epsilon(t)=Cs_0\cos\omega t+Ds_0\sin\omega t, \tag{39}$$

where C and D are known functions of ω. A, the real part of the admittance function, is equal to $D\omega$; and B, the imaginary part, is given by $C\omega$. Thus the impedance function is

$$Z(\omega)=1/(\omega(D+iC)). \tag{40}$$

L. Voigt Model to Impedance Function

It follows immediately from (C) and (K) that the mechanical impedance function, $Z(\omega)=A(\omega)+iB(\omega)$ for a material of known Voigt specification, will be given by the relation:

$$\frac{1}{Z(\omega)}=\frac{1}{\eta}+\sum_j\frac{1}{\eta_j-i\epsilon_j/\omega}+\frac{\omega}{iG_1}. \tag{41}$$

In the case of a continuous distribution of elastic compliance, $J(\tau)$, this becomes:

$$\frac{1}{Z(\omega)}=\frac{1}{\eta}+\int_0^{\infty}\frac{J(\tau)d\tau}{\tau-i/\omega}. \tag{42}$$

M. Maxwell Model to Impedance Function

Similarly, from (D) and (K), it follows that the impedance function for a material of known Maxwell specification will be given by:

$$Z(\omega)=\sum_j \frac{1}{\frac{1}{\eta_j}-\frac{\omega}{iG_j}}, \tag{43}$$

or, in the case of a continuous distribution $G(\tau)$, by:

$$Z(\omega)=\int_0^\infty \frac{G(\tau)d\tau}{\frac{1}{\tau}-\frac{\omega}{i}}. \tag{44}$$

N, O. Voigt or Maxwell Model to Operator Equation

To go from a Voigt model or a Maxwell model to the corresponding operator equation is a little more involved than the reverse transformations (G) and (H). A general method for these conversions has been developed, however, and is described elsewhere.[12] There exists in addition an obvious method which can easily be carried out for models which contain only a small number of elements. To illustrate this simple method, let us consider a Voigt model consisting of two Voigt elements—G_1, η_1, G_2, η_2. The material under consideration is governed by the two differential equations:

$$s=2G_1\epsilon_1+2\eta_1\dot{\epsilon}_1, \tag{45a}$$

$$s=2G_2\epsilon_2+2\eta_2\dot{\epsilon}_2. \tag{45b}$$

The second of these can be rewritten in terms of the total strain, $\epsilon=\epsilon_1+\epsilon_2$ as follows:

$$s=2G_1\epsilon_1+2\eta_1\dot{\epsilon}_1, \tag{46a}$$

$$s=2G_2\epsilon-2G_2\epsilon_1+2\eta_2\dot{\epsilon}-2\eta_2\dot{\epsilon}_1. \tag{46b}$$

Let us now differentiate Eqs. (46a) and (46b), obtaining:

$$\dot{s}=2G_1\dot{\epsilon}_1+2\eta_1\ddot{\epsilon}_1, \tag{47a}$$

$$\dot{s}=2G_2\dot{\epsilon}-2G_2\dot{\epsilon}_1+2\eta_2\ddot{\epsilon}-2\eta_2\ddot{\epsilon}_1. \tag{47b}$$

Equations (45)–(47) can now be combined in such a way as to remove the strain contribution ϵ_1, and its derivatives, by multiplying each equation by the proper factor and adding. The resulting differential equation of the third order involves only the *total* strain and stress and is the desired equivalent operator equation.

P. Creep Function to Voigt Model

Since the creep curve of a viscoelastic material is a known function of the parameters of the specifying Voigt model, it is obviously possible to derive the values of these Voigt parameters from the shape of the creep curve *in those cases where the model consists of a discrete set of Voigt elements with markedly different τ values.* In general, however, this will not be the case. It is still possible to *assume a topological form* for the model, and then compute a reasonable set of parameters for the model. The result is an approximation which may be quite good if enough adjustable parameters are used, but the choice is rather arbitrary. Two different investigators may well obtain entirely different numerical values for the individual elements of their approximating models—even though the over-all behavior of both models is quite similar. This is an obvious disadvantage. A specification, to be most useful, should be unique as well as accurate. It is for this reason that method (U) will be advanced.

Q. Relaxation Function to Maxwell Model

The construction of a Maxwell model from a known relaxation curve is precisely equivalent to the construction of a Voigt model from a known creep curve. Method (V) will be advanced to handle this problem in an accurate, unique fashion.

R. Impedance Function to Operator Equation

Since the impedance function is related to the parameters of the operator equation by Eq. (35), it is possible in principle to derive the operator equation by factoring the known impedance function into a ratio of two polynomials in $(i\omega)$. In practice, using an empirical $Z(\omega)$, this will usually be impracticable. The authors see no simple means for this conversion.

S. Creep Function to Maxwell Model

Simha[13] has shown that the distribution function $G(\tau)$, which defines a continuous Maxwell model, can be derived from the creep curve of the material. The method, which is based upon the use of Fourier transforms, is rather involved and will not be described here. Instead, a much simpler, although approximate, method will be developed in the next section.

[13] R. Simha, J. App. Phys. 13, 201 (1942).

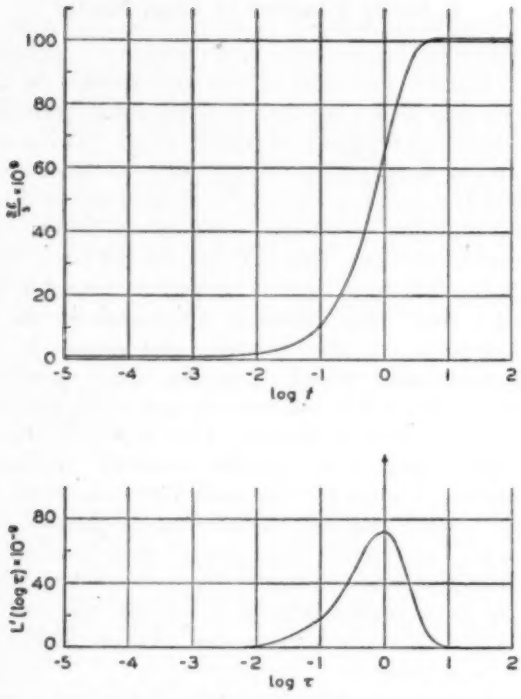

FIG. 6. A creep curve and its distributions of compliance.

T. Relaxation Function to Voigt Model

Because of the symmetry of the relationships among the methods of representation, the problem of deriving an equivalent Voigt model from a relaxation function is mathematically very similar to that of deriving a Maxwell model from a creep function. The method of Simha, with slight alterations, could be used for an exact transformation; an approximate method will be developed in the next section.

III. SOME NEW APPROXIMATION METHODS

The conversions described in the previous section were exact. This advantage, however, is offset by the fact that they involve complicated computations for materials whose viscoelastic models contain a large number of elements. For this reason a set of *approximation* methods will be developed in this section.

U. Creep Curve to Voigt Model

A creep curve shows how strain varies with time at constant stress.[14] The strain at any time will be made up of two parts—elastic deformation and flow. If plotted as 2ϵ *versus* $\log t$, the (*retarded*) *elastic part* of the strain will yield a sigmoidal curve. The total strain can easily be resolved into the two parts, if the creep data cover a wide range of time scale. The viscosity η can be immediately calculated from the flow component. The computation of the viscoelastic spectrum $J(\tau)$ from the sigmoidal elastic creep curve, however, is not so simple. The following method gives an approximation to $J(\tau)$. We will actually derive an approximation to the function $L(\log \tau)$, where $L(\log \tau)d \log \tau = J(\tau)d\tau$.

Plot the elastic part of the creep curve as $(2\epsilon/s)$ *versus* $(\log t)$. Determine the slope of this curve and plot against $(\log \tau)$. This curve which we may call $L'(\log \tau)$ is a fair approximation to the actual distribution function $L(\log \tau)$. $L'(\log \tau)$ fails to reflect sharp discontinuities in L $(\log \tau)$, but reproduces the general contour. If $L(\log \tau)$ is a smooth distribution, extending over a wide range of time scale, then $L'(\log \tau)$ is a very good approximation. If $L(\log \tau)$ contains discontinuities, then $L'(\log \tau)$ is not so good an approximation.

Let us investigate the closeness of the approximation in the most unfavorable case—that of an extremely sharp distribution function. Consider a material whose Voigt model consists of an instantaneous compliance of 10^{-9} cm²/dyne and a retarded elastic compliance of 10^{-7} cm²/dyne whose retardation time is 1 second. (The viscosity of the retarding dashpot is thus 10^7 poise.)

The upper part of Fig. 6 shows the exact creep curve for such a material. The lower part of Fig. 6 shows the derived distribution of compliance $L'(\log \tau)$. This can be compared with the true distribution $L(\log \tau)$, which is the discontinuous function. The area under the hump in $L'(\log \tau)$ is exactly equal to the correct value of J_2, and the maximum of $L'(\log \tau)$ comes at 1 second. Furthermore, the correct value for the instantaneous compliance is obtained. The approximation, however, blurs the sharpness of the distribution. Actually many viscoelastic materials possess distributions which spread over as many as twelve decades of $\log t$.

V. Relaxation Curve to Maxwell Model

In the same way, the distribution of Maxwellian relaxation times can be approximated from the relaxation curve. Plot the relaxation curve as $s/2\epsilon$ *vs.* $\log t$. Determine the slope at each point and plot the negative of the slope *versus* $\log \tau$. The resulting function, which may be called $K'(\log \tau)$, is a good approximation to the true distribution function $K(\log \tau)$—where

$$K(\log \tau)d \log \tau = G(\tau)d\tau.$$

[14] H. Leaderman, Ind. Eng. Chem. **35**, 374 (1943).

W, X. Creep Function to and from Relaxation Function

The creep function, $\phi(t)$, and the relaxation function, $\psi(t)$, are approximately the reciprocals of one another at any time. (*This is not an exact relationship, in spite of the superficial appearance of these two functions.*) The approximate reciprocal relation leads us to a very simple graphical method for going from one function to the other:

Plot the creep function as log (ϕ) *versus* log t. The mirror image of this curve can be obtained by reflecting through the horizontal zero-line, thus replacing every ordinate by its negative. The reflected curve will be an approximation to the relaxation function plotted as log (ψ) *versus* log t. This method obviously works equally well in both directions.

Let us now investigate the closeness of this approximation for a very concentrated distribution of compliance. We already have the creep curve for a Voigt model with $J_1=10^{-9}$, $J_2=10^{-7}$, $\eta_2=10^7$. We can also compute, exactly, the relaxation function for this model. This is shown in the upper part of Fig. 7 (solid curve). Using the known creep curve and the approximation method here advanced, we obtain the approximate relaxation function shown in the upper part of Fig. 7 (dotted line).

It may also be of interest to see how closely this method allows an estimation of the distribution of Maxwellian relaxation times, $K(\log\tau)$, from a creep curve. Using the same model as before, we can obtain three different distribution functions:

1. The exact form of the Maxwell model. This is indicated by the infinitely sharp distribution curve (1) in Fig. 7 (lower part).
2. K' (log τ) can be approximated from the exact relaxation function by means of (V).
3. The approximate relaxation function can be approximated from the creep curve by (W), and then (V) can be applied as a second approximate step. The result is shown in Fig. 7 (lower part).

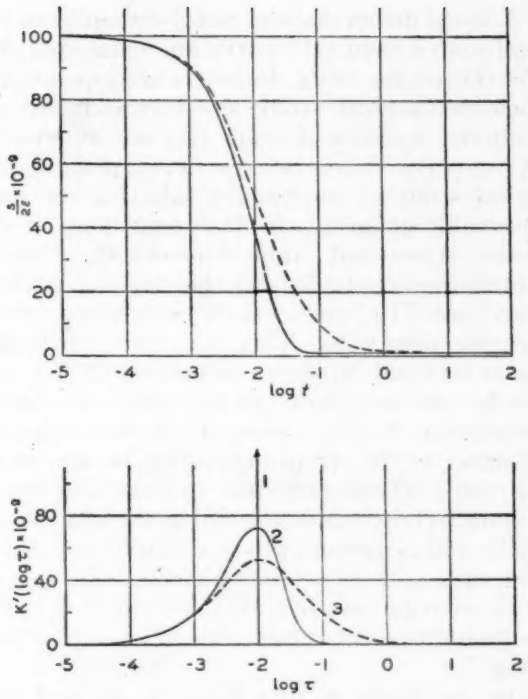

FIG. 7. A relaxation curve and its distribution of Maxwellian relaxation times.

IV. CRITICAL COMPARISON OF THE SEVEN METHODS OF REPRESENTATION

Having discussed, in previous sections, seven methods of representing the mechanical properties of viscoelastic materials, and the important mathematical relationships among these methods, we now proceed to a critical evaluation of the methods.

First of all—as indicated in the discussion—Class I methods would seem to be preferable to the direct experimental functions of Class II from the standpoint of describing the properties of a material. This is because Class I methods are in the nature of *differential* equations which hold for any and all stress and strain sequences, whereas Class II methods are special solutions applying to only one type of experiment. A Maxwellian material, for example, is described in methods (A)–(C) by the differential equation:

$$\dot{\epsilon}=(1/G)\dot{s}+(1/\eta)s. \tag{48}$$

In the special case of a constant strain, ϵ, suddenly imposed at the time $t=0$, the stress follows the specific equation

$$s=G\epsilon e^{-t/\tau}. \tag{49}$$

The difference between Eq. (48) and the relaxation function, $\psi(t)$, is that the former holds for *any* sequence of stresses and strains, while the latter holds only for the one special experiment. For this reason (even though Eq. (49), together with the superposition principle, does *implicitly* define the properties of the material) we will limit the rest of this section to a relative evaluation of Class I methods as *specifications* of properties, followed by a relative evaluation of Class II methods as *experimental* tools.

A sharp distinction can now be made between methods (A) and (B) on the one hand, and (C) and (D) on the other. From the standpoint of a phenomenological study of the mechanics of continua, methods (C) and (D) are superior to (A) and (B). This is because in any phenomenological study of mechanical behavior the only observable quantities are the components of the (total) stress and strain tensors; the "strain contributions" of (A) and the "stress contributions" of (B) are, strictly speaking, unobservable quantities. This very fact, however, makes (A) and (B) more suitable than (C) and (D) for the formulation of molecular theories of deformation.[15–18] In a molecular study, each contribution to the strain may often be identified with some specific molecular process; and hence the strain contributions as well as the total strain can be said to possess a physical significance. Likewise, some authors[19–21] have identified the various stress contributions of a Maxwell model with individual "molecular mechanisms of supporting stress."

Our conclusion is that methods (A) and (B) are more suitable than (C) and (D) for a study of the molecular mechanisms of viscoelastic deformation, but that within the discipline of the mechanics of continua (C) and (D) are to be preferred.

From the standpoint of the mechanics of continua, there is a practical as well as an absolute criterion which must be considered. This is the ease with which a method can be used to predict behavior in special cases; i.e., the ease with which the fundamental differential equations can be solved subject to definite constraints. We may compare Class I methods on the basis of this practical criterion, as follows:

Method (A), the Voigt specification, is extremely well adapted to the solution of all problems in which the stress is specified as a function of time, and the resulting strain is desired. The reverse type of problem, in which a known strain is enforced upon the material and the stress is to be calculated, is very difficult to handle by method (A).

The Maxwell specification is the exact inverse. Problems in which the strain is specified can be solved very easily; those in which $s(t)$ is specified are very difficult.

The operator equation (C) is intermediate in character. Problems involving a known $s(t)$ or $\epsilon(t)$ can both be solved readily, although perhaps not quite so simply as with the more applicable of methods (A) and (B).

If the impedance function is known, problems of both types can be solved by the use of Fourier or Laplace transforms.

We have observed that both the Voigt specification (A) and the Maxwell specification (B) are amenable to molecular interpretation. The important question now arises: Can either of these two methods be said to be superior to the other from the standpoint of a molecular theory? This is a highly debatable question the answer to which at the present time would necessarily be based very largely upon the personal preference of the investigator. The authors' personal preference is for the Voigt specification. From this point of view, the molecular interpretation of viscoelastic behavior may be stated roughly as follows: In the presence of a stress a number of different molecular processes take place.[22] Each molecular process represents some distinct mode of molecular rearrangement. Each process has its own characteristic time dependence and thus results in a strain contribution which is a characteristic function of the time. The total strain in the sample is the sum of the contributions from all the different molecular mechanisms of response to stress. Using this formulation of the problem, a specific molecular theory of the viscoelastic response of one type of material—the amorphous linear high polymer—has been developed.[15]

In terms of the Maxwell specification of properties, the molecular description[23,16] of viscoelastic behavior can be stated in the following manner: The stress on a given cross section of the material is supported by interatomic or intermolecular forces exerted upon the atoms and molecules lying in the cross section. These forces may be given the generic name of "bonds" connecting the material on the two sides of the cross section. Each of these bonds will have its own stiffness and its own characteristic rate of relaxation. A distribution of elastic relaxation

[15] T. Alfrey, J. Chem. Phys. **12**, 374 (1944).
[16] A. V. Tobolsky and R. D. Andrews, J. Chem. Phys. **13**, 1 (1945).
[17] R. F. Tuckett, Trans. Faraday Soc. **38**, 310 (1942).
[18] R. F. Tuckett, Trans. Faraday Soc. **40**, 448 (1944).
[19] R. Simha, J. Phys. Chem. **47**, 348 (1943).
[20] R. Simha, Ann. N. Y. Acad. Sci. **44**, 297 (1943).
[21] A. V. Tobolsky, R. E. Powell, and H. Eyring, *The Chemistry of Large Molecules* (Interscience Publishers, Inc., New York, 1943).
[22] S. Bresler and J. Frenkel, Acta Physicochimica, **11**, 485 (1939).
[23] A. V. Tobolsky, I. B. Prettyman, and J. H. Dillon, J. App. Phys. **15**, 380 (1944).

times is thus correlated on the molecular scale with a distribution in the relaxation times of the interparticle bonds.

The experimental evaluation of the viscoelastic properties of a material can be conveniently divided into two parts—the relatively fast response (say faster than one second) and the relatively slow response (say slower than one second). In principle, either method (E), (F), or (G) can be used to evaluate both aspects of the viscoelastic behavior. In practice, this is not true.

For the rapid part, determination of the steady-state response to sinusoidal stresses (*as a function of frequency*) is the most convenient method. Creep and relaxation studies of the response which occurs in a time scale of 0.01 or 0.001 second are rather impracticable.

On the other hand, creep and relaxation tests are much better suited to the evaluation of the *slow* response than are vibration tests. The dividing line of 1 second, above, is purely arbitrary, and there is no reason why the range of applicability of the "slow" and "fast" methods cannot be made to overlap.

The most promising experimental approach would seem to be the following: Map out the fast response of the material by means of sinusoidal stresses. Then map out the slow response by a creep curve or a relaxation curve. Finally, synthesize the two sets of data and express in terms of one of the fundamental methods of representation—e.g. a Voigt model. The resulting specification of the viscoelastic properties will cover the entire range of time scale—from the fastest frequency of the sinusoidal tests to the total duration of the creep curve.

A Double-Yoke Plate Permeameter

WILLIAM F. BROWN, JR., AND SIDNEY M. RUBENS
Naval Ordnance Laboratory, Washington, D. C.

(Received June 25, 1945)

A double-yoke permeameter, with which both normal and ideal permeabilities of moderately large steel plates can be measured, is described. The technique for making these measurements is also given. The permeameter is suitable for measuring plates several feet long, up to a foot wide, and from $\frac{1}{8}''$ to $1\frac{1}{2}''$ thick; two approximately identical test plates are required.

IT is sometimes necessary to measure the magnetic properties of iron or steel plates $\frac{1}{8}''$ to $1\frac{1}{2}''$ thick, several feet long, and up to 1′ wide. Although ring specimens or specimens suitable for use in standard permeameters may be machined from such plates, the procedure is laborious, destroys the plate, and in the case of soft materials affects the magnetic properties to an unknown degree.

The double-yoke permeameter, illustrated in Fig. 1, was designed to make such measurements directly on plates. This device is a magnetic yoke in which a pair of test plates, both of the same kind of steel, form part of the magnetic circuit. The two test plates, which may be as long as desired and up to one foot wide, are mounted in the yoke so that over a considerable part of their length the plates are magnetized uniformly and in the same direction. This makes it possible to measure both B and H directly with suitable test coils. When the average permeability of large specimens is measured in this way, three disadvantages usually present in measurements on ring specimens of the usual size (outside diameter $\leq 3''$) are obviated: inhomogeneity over short distances, anisotropy, and uncertainty of state of mechanical strain.

The following terminology is convenient in describing the different states of magnetization of a specimen and the methods of producing them. To erase previous history, a specimen is subjected to a series of applied fields of alternating sign and gradually diminishing magnitude; during this procedure, an additional constant field may be superposed. The state of magnetization produced by this procedure is called the "ideal" state; and we shall call the procedure for producing it "idealization," the field used to erase previous history the "idealizing" field, and the superposed constant field the "parameter" field. The idealizing field must begin with a high enough value to produce a result independent of the previous magnetic state, and it must decrease by small enough steps so that if they were to be

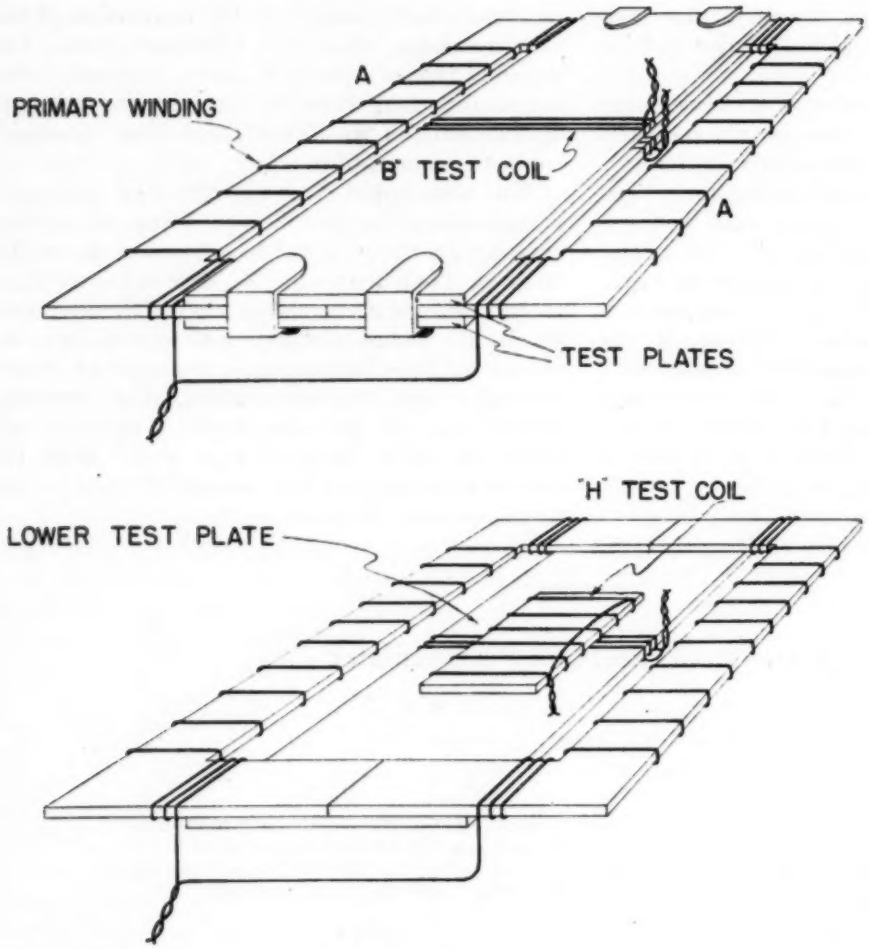

FIG. 1. Construction of the double-yoke plate permeameter. Above, assembled permeameter with both test plates in place. Below, permeameter with top test plate removed to show "*H*" test coil.

made still smaller, no change in the final magnetization value would result. Idealization in zero-parameter field produces zero magnetization and is called "demagnetization." "Normal" magnetization is produced by first demagnetizing the specimen and then applying a field of fixed magnitude alternately in opposite directions until a cyclic state is attained.

The double-yoke plate permeameter was designed primarily to measure normal permeability at low fields, but it may also be used to measure ideal permeability.

DESIGN AND OPERATION

The construction of the permeameter is shown in Fig. 1. The yoke consists of two U-shaped pieces of mild steel 3 feet long which have been sawed from a single plate $\frac{1}{2}$ inch thick. The ends of these yoke pieces are milled smooth on each face so that they make good contact with the two test plates. The middle section *A* of each of the U-shaped pieces is wound uniformly with two electrically separate windings, which provide the primary flux in the yoke. Each winding consists of one layer of number 18 enameled wire. One of these windings may be used to provide a steady field; the other furnishes a field which is varied as may be required for demagnetizing, idealizing, or measuring. The yoke pieces are clamped together with the two test plates. The test plates are 3 feet long, are from 4 to 12 inches wide, and may be of any desired thickness.

Several turns of fine wire are wound around each test plate. These serve as '*B*' test coils. *H* is measured with a long flat rectangular coil wound on a Bakelite form. For plates 10 inches wide or more, an '*H*' coil 12 inches long, 6 inches wide, and $\frac{3}{8}$ inches thick is used to measure fields up to 1 oersted. Its effective *NA* (turns times cross-sectional area) is 4.52×10^4 turns cm^2. This coil lies in the space between the two test plates with its longitudinal axis parallel to the long dimension of the plates. For narrower plates, an '*H*'

coil 2 inches wide having about the same effective *NA* has been used.

The yoke is assembled as shown in Fig. 1. The primary windings of the two sides of the yoke are connected so that the flux from the two sides passes through the test plates in the same direction. The small windings on the short legs of the U-pieces each consist of 100 turns of number 18 enameled wire wound on insulating forms. These can be moved along the short legs if necessary to compensate for flux leakage and improve the uniformity of the flux distribution over the central 12 inches of the test plate. The relative flux distribution obtained in a medium steel test plate $\frac{3}{4}$ inch thick and 12 inches wide is shown in Fig. 2. This flux distribution was measured with a '*B*' test coil whose distance from the center of the plate could be varied.

In the measurements made on test plates of structural steels from 4 to 12 inches wide, no special precautions were required to obtain a sufficiently uniform flux distribution in the test plates. The plates were sawed to the desired size and filed clean where contact was to be made with the yoke pieces. They were clamped to the yoke as tightly as possible with brass *C*-clamps. For the 12-inch wide plates, the four small compensating coils just fitted between the long legs of the yoke and the test plate, so that their positions could not be adjusted. Nevertheless sufficiently uniform flux distributions in test plates of various thicknesses and permeabilities

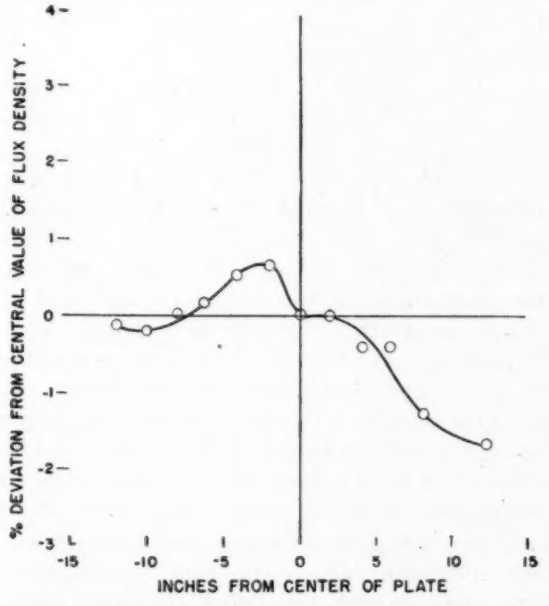

FIG. 2. Test of uniformity of magnetization of test plates: flux distribution in $\frac{3}{4}$" medium steel test plate.

TABLE I. Uniformity of flux distribution in test plates.

Distance from center of plate (inches)	% deviation from flux at center of the test plate		
	Medium steel $\frac{3}{4}$"×11.5"×36"	Medium steel $\frac{1}{4}$"×4"×36"	High tensile steel $\frac{1}{4}$"×12"×36"
6	0.8	0.5	−3.0
4	0.8	0.5	−2.2
2	0.5	0.5	−1.0
0	0.0	0.0	0.0
−2	0.0	0.0	0.3
−4	0.5	0.0	1.4
−6	0.3	0.0	2.2
Smallest detectable fractional difference	1/600	1/200	1/700

were obtained by simply connecting the compensating coils in series with the main windings. Sample data on the flux distributions in 3 different test plates are given in Table I. The smallest detectable difference in fluxmeter readings, expressed as a fraction of the reading, is given at the bottom of the table. If this uniformity had not been so easily obtained with the wide plates, it would have been necessary to adjust the flux distribution by adjusting the currents in the compensating windings.

In order to measure the flux density distribution across a test plate at the center, small holes were drilled through a medium steel plate of dimensions $\frac{1}{4}$"×12"×36" so that three small '*B*' coils could be wound through them. These coils were coplanar and were each 2 inches wide. One was at the center of the plate, and the centers of the other two were +3 and −5 inches from the center of the plate. *B* as measured by each of

TABLE II. Variation in *B* across a test plate at the center.

Position of '*B*' coil	*B*(gausses)
At center of plate	141
3 inches from center	132
−5 inches from center	138
Around entire plate	135

these coils and by a test coil wound around the entire plate at the center under a given state of normal magnetization are listed in Table II. The average of the three small *B* coil measurements is 137, which differs from the value obtained by the coil surrounding the entire plate by less than 2 percent.

The uniformity of *H* in the region between the plates was tested by measuring with the 12" long '*H*' coil at several positions. The coil was moved both across the test plates and along them. The flux through the '*H*' coil was constant

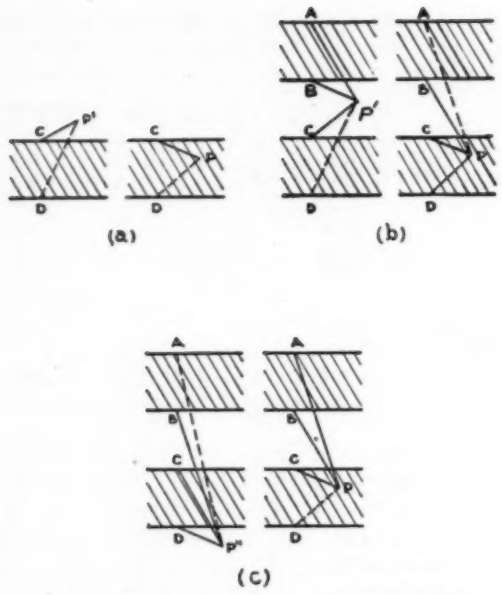

FIG. 3. Contributions of near poles to longitudinal magnetizing force H at a point P in lower test plate, to longitudinal magnetizing force H' at a point P' above it, and to longitudinal magnetizing force H'' at a point P' below it. (a) Comparison of H' and H: lower plate only present. (b) Comparison of H' and H: both plates present. (c) Comparison of H'' and H. For identical pole distributions on equidistant surfaces, the contributions at P from points connected to it by solid lines are equal to the contributions at P' or P'' from points connected to it by solid lines; the contributions corresponding to dashed lines are unequal.

to within ½ percent when the center of the coil was within 6 inches of the center of the plates longitudinally and 3 inches of the center transversely. The 'H' coil was also placed on top of the top plate, at the center, and the value of H was found to be a percent or two lower than in the space between the plates.

The following procedure is used to measure normal permeability. The plates are demagnetized, the desired magnetizing field is applied, and the plates are brought into a cyclic state. Then the magnetizing current is reversed; B and H are measured by observing the deflections of suitable fluxmeters connected to the 'B' and 'H' coils. Directly $B/H=\mu$.

VALIDITY OF THE H MEASUREMENT

With any specimen which is at least approximately uniformly magnetized over part of its length, the measurement of B presents no difficulties; the more difficult problem is the measurement of H. Except in the case of ring specimens and ellipsoids, for which H can be calculated, it is necessary to rely on the continuity of the longitudinal component of **H** across the surface of the specimen. But the quantity actually measured is the average value, H'_{av}, of this component over a finite region adjacent to the specimen, and in many cases this average may differ by 100 percent or more from the average value, H_{av}, over the cross section of the specimen.

In the present permeameter, suppose first that only the lower test plate is in place. The magnetizing force **H** at a point either inside or outside the plate, and near its center, may be analyzed on the basis of the sources into three parts: (1) that directly due to the windings; (2) that due to poles distant, say, more than 5″ from the point in question; (3) that due to poles close to this point. Contributions (1) and (2) are approximately the same in the plate and in the region above it but within ½″ of its top surface. It is contribution (3) that is likely to have serious effects. As is indicated in Fig. 3(a), a pole at a point C on the upper surface of the plate contributes equally to the longitudinal magnetizing force H' at a point P' above the top surface of the plate and to the longitudinal magnetizing force H at a point P an equal distance below it; but a pole at a point D on the lower surface contributes less to H' than to H. Hence if the contribution of local poles is appreciable, H'_{av} may differ significantly from H_{av}. (The contribution of the volume pole density is neglected in this discussion, since it is usually small compared with the contribution of surface poles.)

Addition of the upper test plate improves this situation considerably. An ideally simple case, in which the pole distributions on the four surfaces are identical and the distance between the plates is equal to their thickness, is shown in Fig. 3(b). (The former condition will be closely approximated, if contribution (3) is small compared with contributions (1) and (2).) The contribution at P' from A, B, and C is equal to the contribution at P from B, C, and D; the difference between H' and H is due entirely to the inequality of the distances DP' and AP. Since the contributions of such points as D to H' and of such points as A to H are a small part of the whole contribution of local poles, which in turn is only a fraction of the magnetizing force, the error in equating H'_{av} to H_{av} should be very small under these ideal conditions. In the actual permeameter these conditions are not fully realized; but the error can be estimated by also measuring H''_{av}, the average magnetizing force over an area just below the lower plate or above the upper one. Fig. 3(c) shows that the

difference between H'' and H is due largely to the difference between the contributions of A to H'' and of D to H; these contributions differ by an appreciable fraction of the whole contribution of local poles. Therefore an agreement of H' and H'' to within 2 percent is strong evidence that the contribution of local poles is a small fraction of the magnetizing force, that the deviations from ideal conditions will therefore affect only a small part of the magnetizing force, and that H' and H are equal to within considerably less than 2 percent.

These arguments do not apply to double-yoke permeameters in which the specimens are bars with widths and thicknesses approximately equal. Therefore this plate permeameter appears to have advantages over such permeameters as far as reliability of the H value is concerned, especially since the H'' measurement affords a method of estimating this reliability. For best results, the flux should be made as nearly uniform as possible by adjustment of the auxiliary coil positions and currents; the H' measurement should be checked by a measurement of H''; and the pair of test plates used for the final measurements should be selected ones that give B and H'' values very nearly equal for the two plates.

COMPARISON WITH RING SPECIMEN DATA

To check the reliability of the normal-permeability values obtained with the double-yoke permeameter, a ring specimen 4 inches in diameter was cut from a steel plate one-half inch thick, which had previously been measured in the double-yoke. The ring specimen was wound and measured. The results of both the ring and the double-yoke measurements appear in Fig. 4, in which μ is plotted against H. Although the points agree to within 1 percent, the reliability of the plate permeameter is not necessarily this good. The ring measures an average permeability for all directions of the magnetizing force in the plane of the sample, whereas the plate permeameter measures the longitudinal permeability of the plates; therefore if both methods were perfect, they would give different results for anisotropic materials. The difference for this type of plate is estimated to be about 6 percent. In the results of Fig. 4 this difference is evidently compensated either by other differences between the quantities being measured, such as differences in the space-average permeabilities due to inhomogeneity, or by errors in one or both

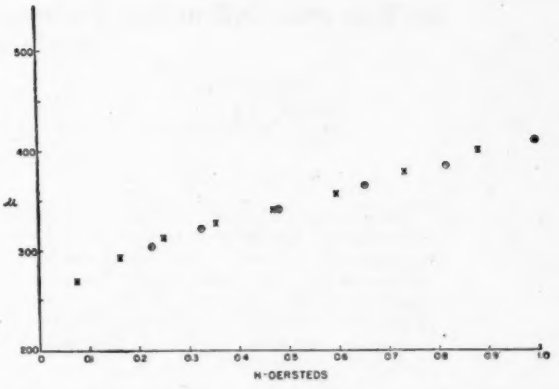

FIG. 4. Comparison of ring specimen and double yoke measurements: variation of permeability with magnetizing force for a sample of medium steel. Crosses, double yoke; circles, ring specimen.

methods. It is probably safe to conclude that the plate permeameter measurements are reliable to within 5 percent.

OTHER APPLICATIONS

The double-yoke permeameter may also be used, with somewhat less precision, to measure the ideal permeability of one of a pair of plates. A steady parameter field is applied by one of the primary yoke windings. The other winding applies the idealizing field. When the idealization is completed, the 'H' coil is rotated about an axis normal to the plates. (The 'H' coil was mounted on a circular cardboard disk which lay on the lower test plate, and the disk could be readily rotated by hand.) The resulting fluxmeter deflection gives H. The clamps are now removed. The top plate is gently lifted off the yoke pieces and the 'B' coil is slipped off the plate. The resulting fluxmeter deflection is proportional to B. In this way, also, states resulting from special procedures, such as incomplete idealization (in which the initial field is not high enough to erase previous history entirely, or the steps in reduction of the field are too large to result in a truly ideal state), may be studied. A small 'H' coil placed on top of the upper plate is convenient for approximate measurement of the idealizing field.

B and H values on a hysteresis loop may be measured by an obvious modification of the technique for obtaining normal magnetization data.

The authors gratefully acknowledge the valuable suggestions given by Mr. J. H. Sweer and the assistance rendered by Mr. J. R. Wright in constructing and testing this device.

An Electron Diffraction Camera for the Study of High Temperature Surface Reactions

EARL A. GULBRANSEN
Westinghouse Research Laboratories, Pittsburgh, Pennsylvania

(Received June 28, 1945)

An electron diffraction camera is described for the study of physical and chemical reactions occurring on the surfaces of metals and alloys *in situ* at elevated temperatures. The specimen is mounted in a metal block and shielded from the surrounding cold surfaces by a series of metal shields. The metal block contains an internal electrical heater which allows the temperature of the specimen to be accurately controlled at temperatures of 800°C. Gas purification chains are included for the preparation of pure oxygen, hydrogen, etc. Several typical electron diffraction reflection studies of the oxidation of iron and stainless steel are included.

INTRODUCTION

THE study of the structure of the reaction product resulting from chemical and physical reactions occurring on surfaces is a problem of increasing importance and one involving many difficulties. Electron diffraction methods have been applied recently with considerable success.

During the past several years we have undertaken a program for the study of the kinetics of formation of oxide and other films on metals. In attempting to interpret the data thus obtained it was found necessary to study the structure of films by electron diffraction methods.

For a complete interpretation it was felt that the specimens of metal should be studied in the electron diffraction camera itself and at given temperature and gas atmospheres. Thus, for example, the specimens could be oxidized in air at 600°C and the diffraction picture taken after evacuating at the same temperature. This would assure a minimum structural change and would allow a series of consecutive pictures to be taken without exposing the sample to the outside atmosphere or allowing the specimen to cool. The effect of heat and gas treatments as well as other processes could be studied *in situ*.

An electron diffraction camera for this type of study requires (1) a furnace capable of being controlled at temperatures up to 800°C, (2) a series of attached gas chains for purifying the gases of interest, i.e., O_2, air, H_2, etc., (3) a rapid pumping system, a pressure measuring system and a valve system for isolating the camera from the pumps, and (4) a film or plate chamber in which 20 to 30 diffraction pictures could be obtained before changing the film.

Two cameras have been described which include specimen holders capable of being heated to high temperatures. The first is the camera described by Yearian and Howe[1] which included a specimen holder for heating the specimens to 600°C and which also allowed the specimens to be cooled to liquid air temperatures. The second is the camera described by Jackson and Quarrell.[2] A carbon arc was used for heating the end of a tube on which the metal specimen was mounted. The temperature of the surface was determined by a thermocouple. A small quantity of gas could be admitted and the reaction product studied *in situ* at the temperature desired.

APPARATUS

The essential features of our camera will be here briefly discussed. A schematic diagram of

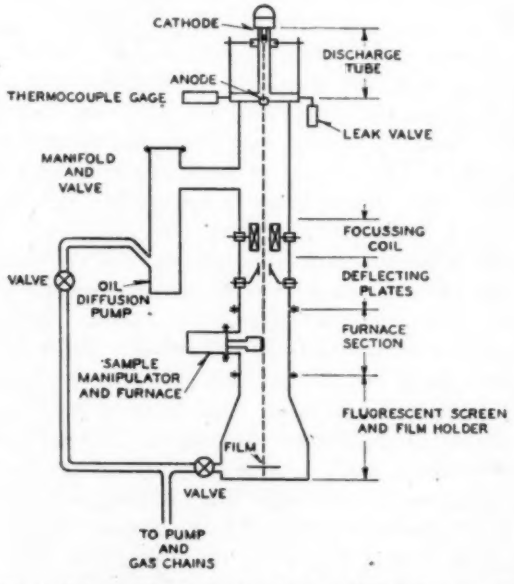

FIG. 1. Schematic diagram, electron diffraction camera.

[1] H. J. Yearian and J. D. Howe, Rev. Sci. Inst. **7**, 26 (1936).
[2] Jackson and Quarrell, Proc. Phys. Soc. London, **51**, 237–43 (1939).

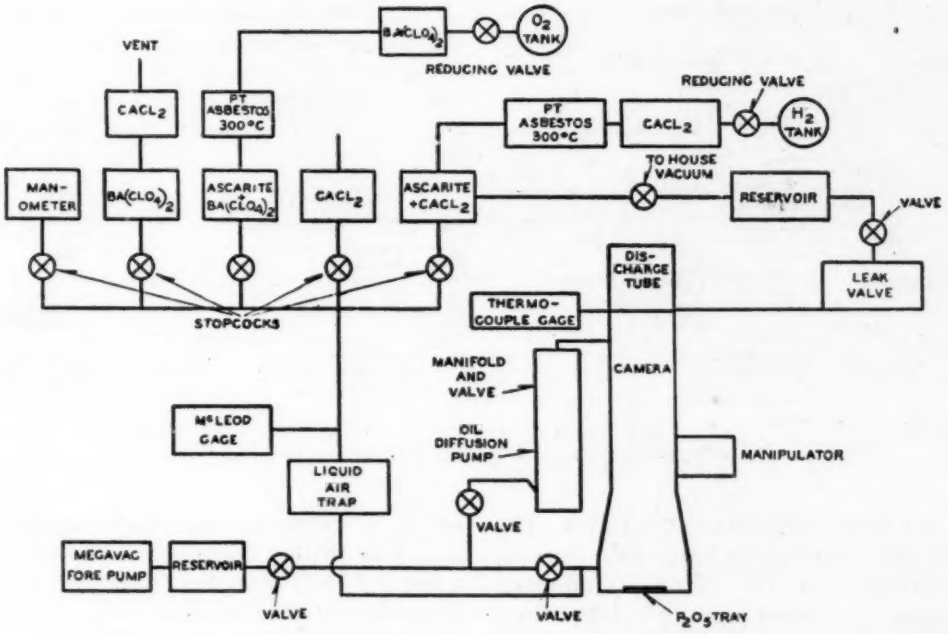

FIG. 2. Preparation chains, pumping, and leak system.

the camera is shown in Fig. 1. A cold cathode source is used to obtain the electrons. Hydrogen gas is used in the discharge tube, the pressure being regulated by varying the pressure of gas behind a porous porcelain leak. For a definite voltage, the current in the discharge tube is dependent only on the gas pressure. Using a voltage of 47.5 kv and a current of 50 microamperes in the discharge tube, exposures of 5 to 10 seconds are necessary with a 0.004-inch anode aperture. The beam current in the camera section is of the order of 10^{-9} ampere. An axial symmetrical magnetic field is used for focusing the electrons on the fluorescent screen and film.

Two sets of deflecting plates are used for moving the beam and, in combination with a movement of the specimen holder, the angle of incidence of the electrons relative to the specimen can readily be changed.

The film chamber is designed for use with roll film. Twenty to thirty exposures can be taken without reloading the chamber. Above the film a fluorescent screen is placed which acts as a shutter to control the exposure time. This is manipulated by a lever and a sylphon bellows.

A metal oil diffusion pump having a speed of 250 l/sec. is used in combination with a Megavac fore pump to obtain the high vacuums. (See Fig. 2.) A three-inch valve is placed in the manifold between the camera and the diffusion pump. The system can also be pumped through an auxillary line direct to the fore pump. A vacuum of 10^{-4} to 10^{-5} mm of Hg is maintained in the main section of the camera while a pressure of about 0.01 mm Hg is maintained in the gas discharge tube.

The pressure in the system can be measured either by a manometer, by a thermocouple gage, or by a McLeod gage. The arrangement of the pressure measuring system, gas preparation chain, and the vacuum system is shown in Fig. 3. Hydrogen gas is purified by (1) drying over $CaCl_2$, (2) removal of O_2 by reaction with H_2 over platinized asbestos, and (3) removal of H_2O and acid constituents by means of $CaCl_2$ and ascarite. Oxygen is purified by (1) drying over $Ba(ClO_4)_2$, (2) conversion of H_2, CO, and other combustibles to H_2O, CO_2, etc., by the use of platinized asbestos, and (3) removal of H_2O, CO_2, etc., over $Ba(ClO_4)_2$ and ascarite.

To facilitate the removal of water and other volatile materials from the film a small tray of P_2O_5 is placed below the film in the camera. The film is partially dried before using by pumping in a separate vessel. A special process 6-122 roll film with non-curling properties is used in the camera although Verichrome film has been used during the emergency. The high voltage is supplied by a conventional type of rectifier, the input of the high voltage transformer being con-

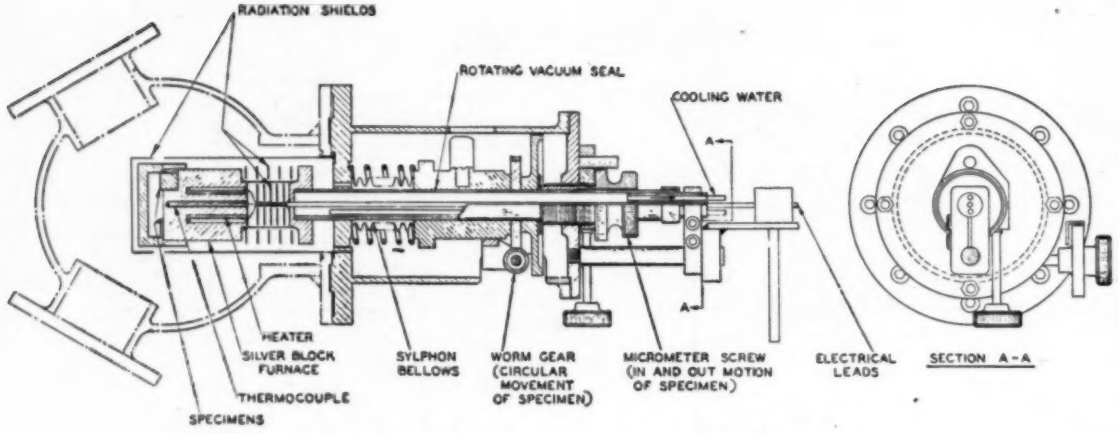

FIG. 3. Electron diffraction camera furnace and manipulator.

trolled by a voltage regulator to $\pm\frac{1}{2}$ volt. The ripple at 50-microampere discharge tube current is less than 0.01 percent. The voltage is calibrated over the range by measuring the voltage drop across a 10,000-ohm resistor in series with a 1000-megohm resistor. The latter is sealed in a glass tube and kept in a nitrogen atmosphere. This external calibration is compared to a calibration based on the diffraction lines of evaporated aluminum or evaporated gold. Over a period of three years the high voltage has been found to be constant to better than $\frac{1}{2}$ or 1 percent.

The specimen holder, furnace, and manipulator are shown in Fig. 3. Two motions are possible. The first motion is a rotation so that several specimens can be examined successively. The second motion is an in and out motion perpendicular to the beam. This allows a change of the incident angle for reflection studies. The rotation is accomplished by a rotating grease vacuum seal and can be set to an accuracy of 1/10 of a degree. The in and out motion operates with a micrometer screw thread on a sylphon bellows and can be set to one ten-thousandths of an inch.

Details of the furnaces are shown in Fig. 4. Two furnaces will be considered. The first is shown in Fig. 4a. It consists of a pure silver block attached to the manipulator head by means of a thin-walled stainless steel section. The silver is furnace brazed to the stainless steel. A silver cap with the appropriate holes for the beam is fitted over the front face and part of the side. Thus the specimens are nearly surrounded by a uniform temperature and the surface temperature is close to the temperature of the furnace. The heater wire is wound on a Lavite core which is fitted into a recess in the silver block. A chromel-alumel thermocouple is placed in a hole drilled centrally in the silver block to a point $\frac{1}{8}$ inch from the front surface. The leads from the thermocouple and heater are taken out through stainless steel tubes. The furnace proper is attached to the head of the manipulator by a vacuum tight fuse wire gasket. The advantage of this arrangement is that the degassing of the wire and Lavite core does not effect the vacuum of the camera. The head of the manipulator is cooled by a stream of water. If it is necessary to operate the heater at extreme temperatures, the inside of the silver block can be evacuated. To reduce heat leakage, radiation shields are placed around the furnace and at other suitable points. The temperature of the surface of a steel sample has been checked by another thermocouple against that given by the block. The temperature of the surface has been found to be within 2 to 14°C of the block temperature at a temperature of 500°C depending on the presence or absence of a gas atmosphere.

A second furnace is shown in Fig. 4b. This is

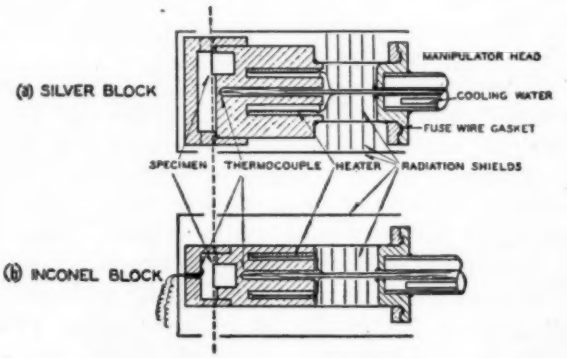

FIG. 4. Furnaces for electron diffraction camera.

designed for one specimen and for use at temperatures up to 800°C. It is constructed of one piece of "Inconel" bar and is mounted on the manipulator head in the same manner as the silver furnace. Because of the poorer heat conductivity of the "Inconel" an additional thermocouple can be mounted on the inside surface of the cap to check the cap surface temperature. A disappearing hot wire type of optical pyrometer is used for measuring the temperature of the furnace cap or specimen. The temperature of the surface is calibrated against the temperature of the block over the temperature range under which the furnace is used.

For transmission type of specimens a special head can be fitted to the manipulator either for screen or for slit mounting of the specimens. In case heat treatment of the transmission specimens is desired, a steel block slightly larger than the usual $\frac{3}{8}$-inch diameter and $\frac{3}{8}$-inch long specimen is fitted into the specimen chamber of the furnace. A hole has been drilled through the block parallel and near to the front face to accommodate the screen and to permit the passage of the incoming and diffracted electrons. This is useful for the heat treatment of evaporated metal films on silica or stripped oxide films.

CALIBRATION

Figure 5 shows schematically the variables involved in the analyses of reflection patterns. The poly-crystalline specimen at a distance L from the film diffracts electrons giving rings of radii R_{hkl}. The undiffracted beam is shown striking the film behind the shadow edge. Bragg's law relating the diffraction angle to wave-length and lattice spacing can be applied to electron diffraction as well as to x-ray diffraction. For small angles Bragg's law simplifies to $R_{hkl}=\lambda L/d_{hkl}$. Here λ is the wave-length and d_{hkl} is the lattice spacing effective in producing the diffraction $_{hkl}$.

The instrument can be calibrated by directly determining the voltage and thus λ or by measuring the diffraction pattern of known materials and calculating the values of λ. Table I shows a comparison of the d_{hkl} of evaporated aluminum, (taken by reflection) calculated from the voltage calibration and those listed in Hanawalt, Rinn, and Frevel's[3] tables. The agreement is good to about 1 percent. This is what is expected because

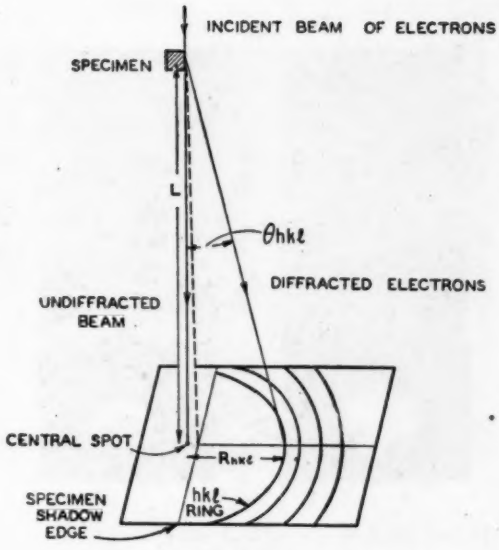

FIG. 5. The reflection method.

of the uncertainty of the plate to specimen distance due to the width of the specimen. The width of the specimen is 1 cm and the specimen to plate distance is 48 cm.

PREPARATION OF SPECIMEN FOR REFLECTION STUDIES

The technique of preparing the surface for study is extremely important since the diffraction of the electrons impinging on to the surface is markedly affected by (1) the flatness of the surface, (2) the area exposed to the beam, (3) the nature and size of the asperities present, and (4) the presence of impurities. According to Thomson and Cochrane[4] the aim is to obtain a clean, flat, and slightly matted surface of the material. For many applications the specimen is ground on various grades of polishing paper using either

TABLE I. Diffraction pattern of evaporated aluminum.

Radius in cm	Electron diffraction		X-ray tables	
	d_{hkl} in A	Intensity	d_{hkl} in A	Intensity
1.09	2.31	Strong	2.33	Strong
1.25	2.015	Strong	2.02	Strong
1.78	1.415	Strong	1.43	Strong
2.07	1.215	Strong	1.219	Strong
2.17	1.16	Weak	1.168	Weak
2.73	0.922	Weak	0.928	Weak
2.81	0.895	Weak	0.905	Weak
3.08	0.820	Weak	0.826	Weak
3.29	0.765	Weak	0.778	Weak

[3] Hanawalt, Rinn and Frevel, Ind. Eng. Chem. Anal. Ed. 10, 457–512 (1938).

[4] Thomson and Cochrane, *Theory and Practice of Electron Diffraction* (The Macmillan Company, New York, 1939).

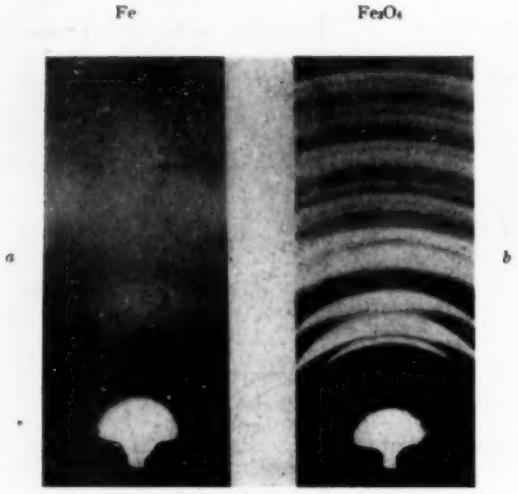

FIG. 6. Oxidation pure iron 300°C at 1.5 cm air. *a*. Vac. at 25°C. *b*. Oxidized 47 min.

benzene, kerosene, or a solution of paraffin in kerosene. These solvents serve as a lubricant and aid in the prevention of oxidation of the metal. After removing the lubricant the specimens can be washed and dried in absolute alcohol. The specimen may be gently etched in order to remove surface oxide films and the surface layer of the metal. However, extreme care must be observed to remove traces of impurities left on the surface.

Recently Vernon and co-workers[5] have studied the effect of polishing on the surface film formed on chromium nickel 18–8 stainless steel by chemical analyses of the oxide film. They find that the higher the degree of polishing the thicker is the resulting oxide film and the greater the change in the chromium to iron ratio in the surface oxide. This work indicates that a high degree of polishing is to be avoided if the surface is to take part in subsequent chemical reaction.

To prepare flat, clean, and abraded surfaces Mr. G. Rose of our laboratories has developed a precision abrader. To minimize oxide film formation and changes in the composition of the surface layer it was felt that the heat generated in the abrasion process should be reduced to a minimum without the use of lubricants. This frictional heat can be partially removed if fresh surface is continually exposed to the metal. This avoids metal to metal contact as a result of the polishing particles picking up the metal from the surface.

[5] Vernon, Wormwell, and Nurse, Metallurgia **31**, 19–22 (1944).

The abrader consists of the following parts: (1) a glass plate over which a piece of polishing paper is tightly stretched, (2) a pair of connected precision ground parallel guide bars which can be shifted at will across the paper in definite strips, (3) a flat plate with suitable grooves for guiding on the parallel bars, (4) a brass chamber mounted in the center and perpendicular to the plate, and (5) a stainless steel piston in which the specimens are mounted. The abrading pressure can be regulated by adding weights to the stainless steel piston. After having been placed in the chamber the piston with specimen can take only two positions relative to the guide bars. This is to facilitate parallel and 90 degree abrading. The plate plus chamber and piston with specimen is pushed back and forth on the guide bars with the specimen in contact with the paper.

The abrading treatment of a particular sample is determined by its hardness. Soft metals require different pressures and a more frequent change of polishing paper. The polishing papers used are 1, 0, 00, 000, and 0000. The last stage of the process is carried out exposing a new section of paper for each stroke and rotating 90 degrees between each stroke. The abrading pressure is lowered during the last stage of the process.

Using the above method we have successfully abraded pure aluminum and pure magnesium in the absence of lubricants.

Following abrasion treatment the specimens can be given any pre-treatment desired before carrying out the experiment in the electron diffraction camera. The specimens are mounted in the instrument with the direction of abrasion at 45 degrees to the electron beam.

TYPICAL RESULTS

Figure 6 shows the oxidation of pure electrolytic iron at 300°C. The specimen was abraded with 1, 0, 00, 000, and 0000 polishing paper. Part (*a*) shows the diffraction picture of the metal taken by reflection at room temperature. Faint iron lines can be seen. Part (*b*) is the diffraction pattern obtained after 47 minutes of oxidation at 1.5 cm of air at 300°C. The pattern indicates Fe_3O_4.

Figure 7 shows the oxidation of pure electrolytic iron at 400°C and the effect of heat treatment on the resulting pattern. Part (*a*) is taken of the metal at room temperature while part (*b*) is taken at 400°C in a vacuum before air is

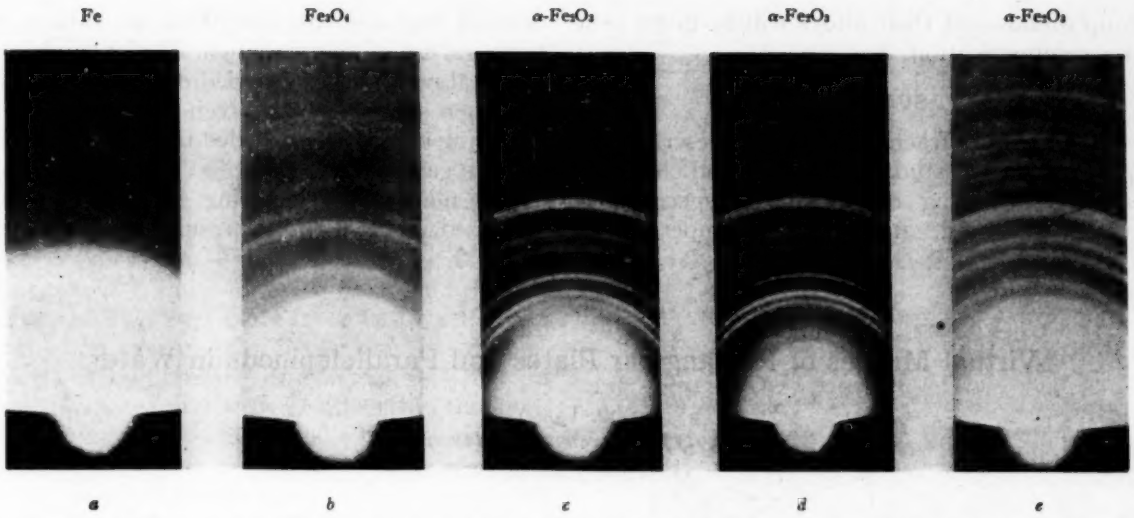

FIG. 7. Oxidation pure iron 400°C 0.1 atmos. air. *a*. Polished No. 1 grit 25°C. *b*. Vac. at 400°C. *c*. Oxidized 60 min. air. *d*. Heated to 500° for 30 min. *e*. Cooled to 25°C overnight.

added. The pattern is Fe_3O_4. Part (c) is the pattern resulting from 60 minutes of oxidation at 400°C. and 0.1 atmos. of air. The pattern indicates α-Fe_2O_3. The effect of temperature is shown in part (*d*) taken after heating to 500°C for 30 minutes. No change in the pattern is noticed. Part (*e*) is the diffraction pattern obtained after cooling to room temperatures. No change in the pattern is found.

Figure 8 shows the oxidation of stainless steel (18–8) at 550°C. Part (*a*) shows the pattern obtained in a vacuum at 25°C. The lines can be fitted with either an iron or a chromium lattice since the lattice constants are similar. Part (*b*) shows a pattern taken after oxidizing at 550°C in 7.5 cm of air for 5 minutes. The structure is either α-Fe_2O_3 or Cr_2O_3, since the lattice constants are very similar for the two oxides. Part (*c*) shows the same structure after continuing the oxidation for 50 minutes. To determine the equilibrium oxide present at 200°C the specimen was cooled to 25°C and then heated at 200°C for 16 hours. Part (*d*) shows the resulting pattern which is the same as that found at 550°C. The specimen was now heated to 600°C. Part (*e*) shows the stability of the resulting oxide at 600°C.

A complete study of the formation and stability of the oxide films on the iron transition

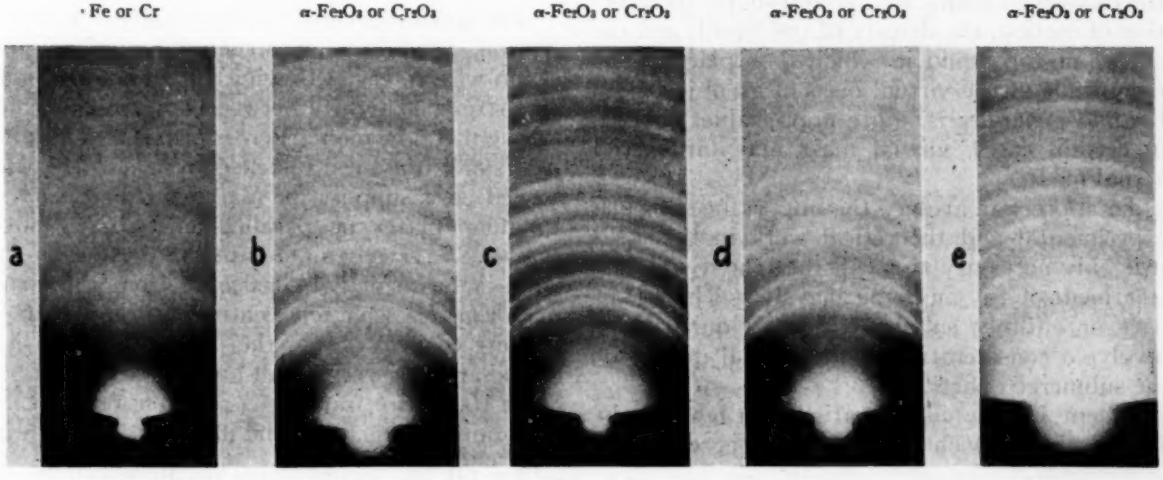

FIG. 8. Oxidation 18–8 stainless steel 550°C 0.1 atmos. air. *a*. Vac. 550°C. *b*. 5 min. air. *c*. 50 min. air. *d*. After cooling to 25°C and heating at 200°C for 16 hrs. (taken at 200°C). *e*. Same as (*d*) but taken at 600°C.

group metals and their alloys will be given in a later communication.

SUMMARY

An electron diffraction camera is described of special use in the study of physical and chemical reactions occurring on metals in special gas atmospheres and at elevated temperatures. Several furnaces are described for raising the temperature of the specimen. The chemical and physical reactions are carried out *in situ* and the specimen studied at the given temperature. Gas purifying chains are provided for the preparation of pure gases.

The method of preparing samples is briefly discussed. Several typical results are given.

Virtual Masses of Rectangular Plates and Parallelepipeds in Water

YEE-TAK YU

*University of Cincinnati, Cincinnati, Ohio**

(Received June 26, 1945)

The following is a report on experimental results obtained for the virtual masses of thin rectangular plates and rectangular parallelepipeds moving in water. Experimental data, together with an empirical equation, showing the manner in which the virtual mass of a rectangular plate varies with dimensions and direction of motion are given. Data and an empirical equation giving the virtual mass of a rectangular parallelepiped as a function of dimensions and direction of motion, for the special case in which the motion is parallel to one face, are also listed.

INTRODUCTION AND SCOPE OF THE WORK

IT is a well-known fact that the force required to accelerate an object in a liquid is greater than the force required to give it the same acceleration in space. This is an effect over and above any viscous or other dissipative force which may be acting. As a result, the effective mass of an object is greater in a liquid than in space. This increase in effective mass is referred to as "virtual mass." The magnitude of the virtual mass of any particular object depends on the following factors: its size and shape, its direction of motion, the density of the liquid, and the extent of the liquid in which it is placed. The magnitude of the virtual mass of an object may be, and generally is, quite appreciable. Indeed, in certain cases, virtual mass may far exceed actual mass.

To appreciate at once the importance of both experimental and theoretical work in this field it is only necessary to recall that the dynamical treatment of any mechanical system, which is in part or entirely submerged in a liquid, must involve a consideration of the virtual masses of the submerged parts.

Except for the consideration of a few simply shaped objects such as a sphere,[1] long cylinder, etc., very little theoretical work has been done. A search through the literature reveals that even less has been accomplished experimentally.

The following is an account of the results of a series of experiments in which the virtual masses of thin rectangular plates and rectangular parallelepipeds in water were determined as a function of linear dimensions and as a function of direction of motion. The work herein reported is a continuation of previous work[2] in this field.

APPARATUS AND EXPERIMENTAL PROCEDURE

The apparatus consisted essentially of a simple torsion pendulum constructed as follows. A piano wire *w*, Fig. 1, having a diameter of 1 mm was stretched tightly between opposite legs of a strong metal frame *F*. A cross-arm *AA* consisting of a small brass rod fastened rigidly to the wire served as a support for a pair of identical rectangular plates or parallelepipeds *BB*. Large screws *SS* served to regulate the tension on the wire. A beam of light reflected from the small mirror *m* to a scale ten to fifteen feet away made it quite easy to obtain the period of oscillation of the arm for very small amplitude of motion. Since the entire frame was portable it was easy to obtain periods in air and in water. (Water was

* Now at Research Laboratory, U. S. Naval Ordnance plant, Indianapolis, Indiana.

[1] Horace Lamb, *Hydrodynamics*.

[2] Yee-Tak Yu, "Virtual Masses and Moments of Inertia of Disks and Cylinders in Various Liquids," J. App. Phys. **13**, 66–69 (1942); "Double Torsion Pendulum in a Liquid," Am. J. Phys. **10**, 152–153 (1942).

the only liquid used since theoretical consideration, as well as previous experimental work, had shown that all virtual masses are directly proportional to the density of the liquid.) The container for the water was a large cylindrical tank 4 feet in diameter and 5 feet deep. The apparatus was submerged to a point where the mirror m was just above the surface of the water.

The experimental procedure was simple. Identical rectangular plates or rectangular parallelepipeds BB were rigidly fastened to either end of the cross-arm of the torsion pendulum. Periods of oscillations, first in air and then in water, were carefully determined by timing a large number of oscillations with an accurate stop watch. These two periods, together with certain other data, to be mentioned presently, were sufficient for computing the virtual mass of either object.

The period of a simple torsion pendulum of the above type in air is given by

$$T_a=2\pi(I_a/K)^{\frac{1}{2}}, \tag{1}$$

and neglecting the damping due to the viscosity (which on the basis of previous experience is quite negligible) the period of this pendulum in water is given by

$$T_w=2\pi(I_w/K)^{\frac{1}{2}}. \tag{2}$$

Where I_a and I_w are the moments of inertia of the cross-arm with objects BB attached in air and in water, respectively, and K is the torsional constant of the wire w. Now let I_a' and I_w' represent the moments of inertia of the cross-arm only in air and in water, respectively. Then the increase in moment of inertia due to a single object B moving in water is given by

$$I=\tfrac{1}{2}[I_w-I_a-(I_w'-I_a')]. \tag{3}$$

This difference we shall refer to as the virtual moment of inertia of the object in the water. Letting L represent the distance from the center of the wire w to the center of the object B and assuming that L is relatively long we can write:

$$I=ML^2 \tag{4}$$

where M is now the *virtual mass* associated with B. Hence, assuming that the torsional constant K is known, Eqs. (1)–(4) give the final result we are seeking. It will be noticed that the same procedure can be used for finding virtual mass regardless of how the objects are oriented on the ends of the cross-arm. Thus this same apparatus and procedure was used as a means of determining the virtual masses of rectangular

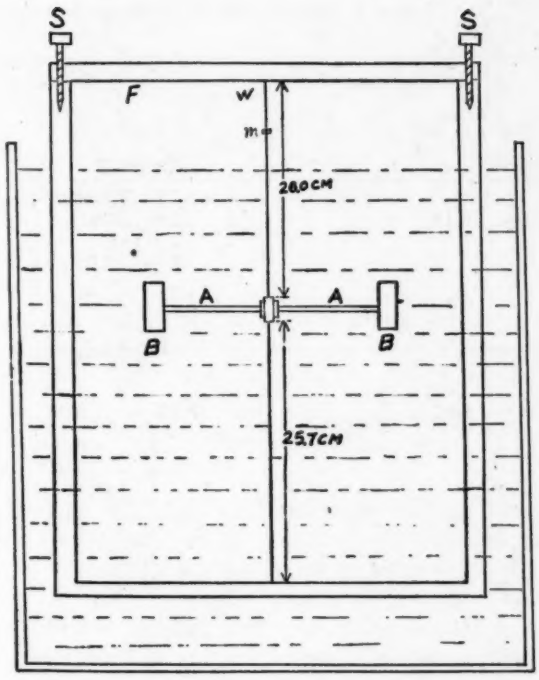

FIG. 1. Diagram of apparatus.

plates and rectangular parallelepipeds as functions not only of linear dimensions but of direction of motion as well.

The constant K was determined experimentally by replacing the cross-arm shown in Fig. 1 with a long uniform brass rod of known moment of inertia I_k. I_k is related to the period of oscillation T_k in air by,

$$K=4\pi^2(I_k/T_k^2).$$

With this and previously given relations it is easy to derive the following convenient working equation,

$$M=\frac{I_k}{2T_k^2L^2}[T_w^2-T_a^2-(T_w'^2-T_a'^2)], \tag{5}$$

where T_w' and T_a' represent the periods of the pendulum in water and in air respectively *without* objects BB. The values of T_k, T_w', and T_a' remain constant. However, it was found that relatively small changes in the tension would change these values slightly. Hence these periods were checked frequently.

Rectangular plates for this work were cut from thin sheet lead. Rectangular parallelepipeds were cast from plaster of paris in a simple mold. After hardening and drying the plaster of paris casts were boiled in paraffin wax for some time

TABLE I. Data for the parallelepipeds moving into water perpendicularly to their faces $2a \times 2b$ ($I_k/T_k^2 = 7.28 \times 10^3$).

$2a \times 2b$ in cm² / T in sec. / L in cm / M in g / $2c$ in cm.	8×5 $M_1 = 133.2$		6×5 $M_1 = 90.4$		4×5 $M_1 = 49.0$	
0.15	$T_a = 3.72$ $T_w = 6.08$ $L = 24.0$ $M = 143.3$ $M_2 = 10.1$		$T_a = 3.32$ $T_w = 5.10$ $L = 24.0$ $M = 96.7$ $M_2 = 6.3$			
0.3	5.19 23.9 146.9 13.7	7.10	4.55 23.9 100.9 10.5	6.08	3.77 23.9 55.4 6.4	4.82
1.0	3.42 23.3 158.4 25.2	5.98	2.92 23.3 108.5 18.0	5.00	25.1 23.3 61.9 13.0	3.98
2.0	4.65 23.3 171.4 38.2	6.90	5.71 23.3 116.8 26.4	3.86	3.30 23.3 65.3 16.3	4.57
3.0	5.28 23.3 177.2 44.0	7.40	4.76 23.4 120.2 29.8	6.40	3.88 23.3 70.0 21.0	5.07
4.0	6.00 23.3 182.7 49.5	7.98	5.43 23.3 126.6 36.2	6.97	4.31 23.3 73.8 24.8	5.47
5.0	6.79 23.3 190.7 57.5	8.66	6.00 23.4 130.0 39.6	7.47	4.97 23.3 75.4 26.4	6.02
6.0	7.65 23.4 194.2 61.0	9.39	6.62 23.3 133.4 43.7	8.00	5.47 23.3 80.1 31.1	6.49
7.0			6.76 23.3 137.0 46.6	8.15		
8.0			7.64 23.3 140.1 49.7	8.92		

to render them impervious to water. A brass nut was cast in the face of each so that it could be screwed directly to the end of the cross-arm.

EXPERIMENTAL RESULTS

A. Virtual Masses of Rectangular Parallelepipeds With Motion Perpendicular to One Pair of Faces

Following the procedure just outlined the virtual masses of three groups of rectangular parallelepipeds were determined. Each group consisted of seven or more pairs of blocks of varying thickness. Each block of a particular group presented the same cross-sectional area (8×5, 6×5, and 4×5 sq. cm, respectively) perpendicular to the direction of motion. Results are given in Table I. Letting $2a$, $2b$, and $2c$ represent the dimensions of a block ($2c$ being the dimension parallel to the direction of motion) the empirical Eq. (6) gives a value of the virtual mass of any one of the blocks in close agreement

with experiment. A brief synopsis of the way in which this equation was arrived at will now be given.

Graphs of Fig. 2 show the result of plotting virtual mass as a function of $2c$ for each of the groups. Consider any one of the curves. It is evident that it can be represented by $M=M_1+M_2$ where M_1 is the value obtained by extrapolating the curve to zero thickness. M_2 is a function of thickness. (For any one curve $2a\times 2b$ is constant but $2c$ varies). A plot of log M_2 against log c for each family of blocks gives the straight lines of Fig. 3. The slope of each is approximately $\frac{1}{2}$. Hence M_2 for any one of the curves, Fig. 2, is represented by $M_2=BC^{\frac{1}{2}}$. A comparison of the values of B, as obtained from each of the straight lines, shows that it is approximately $3.5\rho ab$. The density of the liquid ρ is introduced here because it is known that in all cases virtual mass is directly proportional to this quantity. Thus we write:

$$M=M_1+3.5\rho abc^{\frac{1}{2}}. \tag{6}$$

M_1, the virtual mass for $2c$ approaching zero, undoubtedly depends on $2a$ and $2b$. In order to find how M_1 depends on a and b the virtual mass of a number of lead plates, each having the same thickness (0.3 cm) but varying in area from

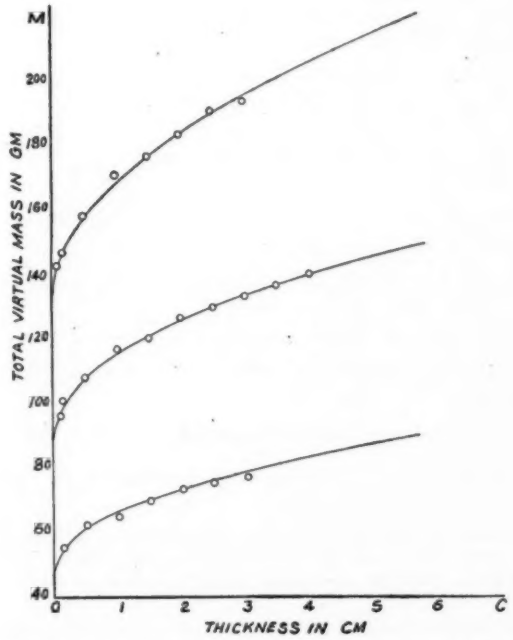

FIG. 2. Virtual masses of rectangular parallelepipeds moving perpendicular to the faces of constant areas (8×5, 6×5, and 4×5 cm^2, respectively) as a function of thickness $2c$.

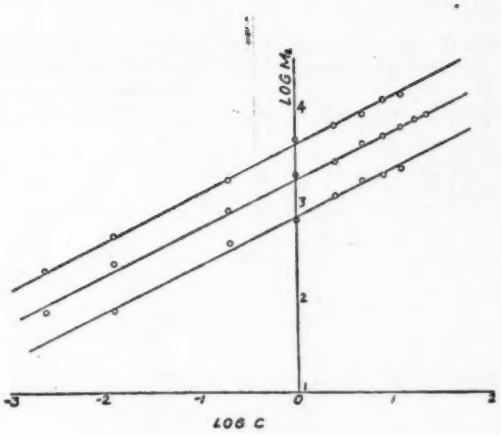

FIG. 3. Plot of log M_2 against log C (Slope = approx. $\frac{1}{2}$).

4×4 to 4×12 sq. cm, was determined. Results are shown in Table II, column 1. (These same data are also an important part of Section B of this paragraph.) Since the plates are so thin, values of M are almost, but not quite, equal to M_1. Assuming the correctness of Eq. (6) a value of M_1 was computed for each experimental value of M. From certain theoretical considerations to be mentioned later, it appeared likely that M_1 should be proportional to $a^2b^2/(a^2+b^2)^{\frac{1}{2}}$. A plot of M_1 against this quantity gives the straight line shown in Fig. 4, the slope of which is 6.3. Hence we finally write as the total virtual mass of a rectangular parallelepiped having dimensions $2a$, $2b$, $2c$ and moving perpendicular to the face $2a\times 2b$,

$$M=\rho\left[\frac{6.3a^2b^2}{(a^2+b^2)^{\frac{1}{2}}}+3.5abc^{\frac{1}{2}}\right]. \tag{7}$$

That this relation gives results in close agreement with experiment can be seen by comparing experimental and computed values given in Table III.

B. Variation of Virtual Mass of a Thin Rectangular Plate with Direction of Motion

It is evident from elementary consideration that the virtual mass of a rectangular plate moving edgewise is less than when moving perpendicular to its face. Hence it is clear that virtual mass must in general depend on the direction of motion. In order to determine this dependence the following procedure was adopted. Exactly similar rectangular plates (the thin lead plates previously used) were rigidly fastened to either end of the cross-arm of the original

TABLE II. The variation of virtual mass of a rectangular plate with direction of motion. (In these data all plates have the same thickness, 0.3 cm. The effective arm L in each case was 23.55 cm. θ designates the angle between the normal to the face and the direction of motion.)

T in sec. θ M exp in g M comp in g $2a \times 2b$ in cm²	0	$\pi/12$	$\pi/6$	$\pi/4$	$\pi/3$	$5\pi/12$	$\pi/2$
	T_a T_w	T_a T_w	T_a T_w	T_a T_w	T_a T_w	T_a T_w	T_a T_w
4×4	3.35 4.21	3.35 4.16	3.35 4.04	3.35 3.84	3.36 3.68	3.36 3.52	3.36 3.47
	M_{exp}	M_{exp}	M_{exp}	M_{exp}	M_{exp}	M_{exp}	M_{exp}
	40.3	37.0	31.0	20.7	12.1	4.6	1.8
	M_{comp}	M_{comp}	M_{comp}	M_{comp}	M_{comp}	M_{comp}	M_{comp}
	41.0	38.3	31.2	21.4	11.8	4.4	1.7
4×6	3.96 5.16	3.97 5.10	3.97 4.97	3.98 4.68	3.98 4.38	3.99 4.20	4.00 4.10
	70.2	64.7	54.8	37.6	19.6	7.4	2.7
	70.8	66.2	53.6	36.4	19.3	6.8	2.1
4×7	4.28 5.64	4.28 5.58	4.29 5.36	4.29 5.07	4.30 4.74	4.30 4.48	4.30 4.40
	87.1	82.5	66.1	45.9	24.4	7.7	2.7
	87.1	81.2	66.0	44.6	23.4	7.7	2.1
4×8	4.56 6.03	4.56 5.97	4.57 5.72	4.57 5.40	4.57 5.10	4.58 5.05	4.58 4.66
	102.5	96.2	76.1	52.6	28.2	9.1	2.9
	102.0	95.2	77.1	52.2	27.9	9.1	2.4
4×12	5.46 7.38	5.47 7.27	5.48 7.00	5.49 6.55	5.50 6.08	5.51 5.74	5.52 5.62
	160.4	148.2	122.9	81.8	41.7	14.2	2.9
	159.4	147.2	120.7	81.2	42.2	15.0	3.0

apparatus (Fig. 1) in such positions that the normal to the face of each was at a definitely known angle θ with respect to its direction of motion. This angle could be set to within an accuracy of about ± 0.5 degree by means of a simple protractor arrangement which we shall not describe. A set of values of virtual masses for rectangules of various dimensions $2a$ and $2b$, but all with the same thickness (0.3 cm), for angles 0, $\pi/12$, $\pi/6$, $\pi/4$, $\pi/3$, $5\pi/12$, and $\pi/2$ radians between the normal to the face and the direction of motion are given in Table II. These rectangles are so thin that they can almost be regarded as having only two dimensions. Plotting experimentally determined values of virtual mass against corresponding angles for each rectangle the curves of Fig. 5 were obtained. In plotting these graphs, values between 0 and 90 degrees are repeated for corresponding angles on the other side of the normal.

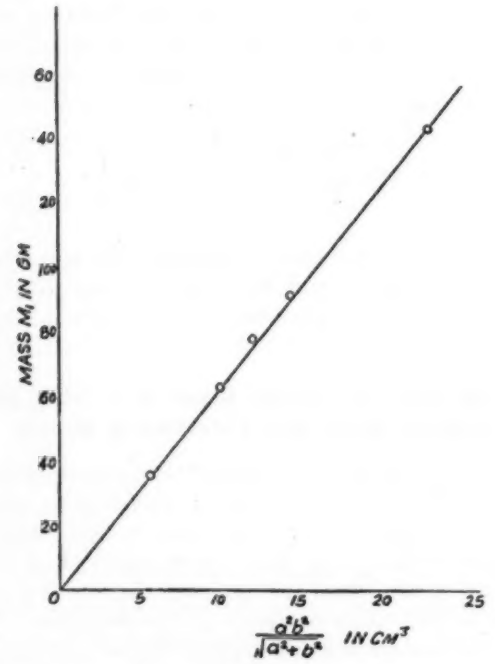

FIG. 4. Graph showing the relation between M_1 and $\frac{a^2b^2}{(a^2+b^2)^{\frac{1}{2}}}$ (where slope = 6.3).

Any one of these curves is represented quite accurately, except for angle near $\pm\pi/2$, by the equation

$$M_\theta = M_0 \cos^2 \theta. \tag{8}$$

That Eq. (8) does not give results in close

TABLE III. Results obtained by computation from Eq. (7) and experimental results with rectangular parallelepipeds moving in water perpendicular to the face $2a \times 2b$.

$2a \times 2b \times 2c$ in cm³	17×9×4	8×4×2	5×5×5
M in g (Eq. 7)	1151	120	104
M in g (exp.)	1148	118	104

agreement with experiment for angle near $\pm\pi/2$ is owing largely to the fact that rectangles have a finite thickness. This fact will be taken account of in the work that follows. From the results just discussed it seems certain that, neglecting thickness, the virtual mass of a rectangle varies as the square of the cosine of the angle between the direction of motion and a normal to the face. Combining this with previous results we write

$$M=\rho\left[\frac{6.3a^2b^2}{(a^2+b^2)^{\frac{1}{2}}}\right]\cos^2\theta. \qquad (9)$$

Note agreement between Eq. (9) and experiment as shown in Table II.

THEORETICAL CONSIDERATIONS

The general theory of the virtual mass of an object moving in a perfect fluid has been completely developed.[3] Unfortunately, however, it is quite difficult to apply the theory to most objects, even though their geometrical shapes may be rather simple. So far as is known, no theoretical work on rectangular plates or parallelepipeds has been reported. However, Horace Lamb does give $M=\pi\rho b^2$ as the virtual mass per unit length of an infinitely long strip of breadth $2b$ when motion is perpendicular to its face. Hence, so far as agreement between theory and experiment is concerned, only two remarks can be made. The factor 6.3 occurring in Eq. (9) was obtained as the slope of the straight line Fig. 4. Now let $a\gg b$ and $\theta=0$. Then neglecting b in the denominator of Eq. (9) and dividing through by $2a$ we have $M=3.15\rho b^2$ which should represent the virtual mass per unit length of a very long strip. But this is exactly the relation given by Lamb. Secondly the $\cos^2\theta$ occurring in Eq. (9) is to be expected from theoretical consideration. One of the boundary conditions over any surface is that the normal component of velocity of the fluid particles in contact with the surface must be equal to $V\cos\theta$. Over a plane surface where θ remains constant $V\cos\theta$ remains constant. Hence, the general expression for kinetic energy T will contain the square of this factor.

[3] Horace Lamb, *Hydrodynamics*, Third edition, Chap. III.

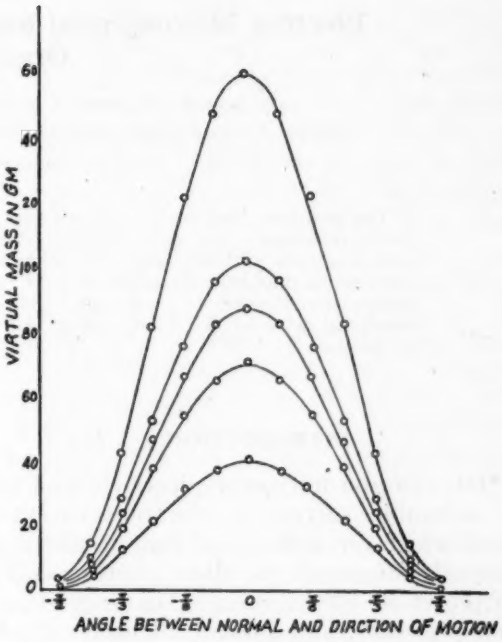

FIG. 5. Virtual mass *versus* angle of motion.

CONCLUSIONS

Insofar as comparison of experimental results with general theoretical considerations can be made, there is qualitative agreement. It is believed that the data herein given are sufficiently accurate to serve as a check on specific theoretical expressions which may later be reported.

The empirical equations give results in close agreement with experiment and it is hoped, therefore, that they will be of immediate use in the field of applied science.

The author wishes to express her thanks to Dr. D. A. Wells for many helpful suggestions and discussions.

Electron Microscopical Replica Techniques for the Study of Organic Surfaces

R. Bowling Barnes, Charles J. Burton, and Robert G. Scott
Stamford Research Laboratories, American Cyanamid Company, Stamford, Connecticut

(Received July 2, 1945)

The problem of examining opaque surfaces in the electron microscope is discussed and it is shown that none of the methods heretofore described is completely satisfactory for the examination of organic surfaces. Two new methods, a low pressure polystyrene-silica technique and a silver-silica double evaporation method, are described and it is demonstrated that they allow observation of organic as well as metallurgical surfaces. Electron micrographs of fiber surfaces, biological material, paint films, ores, and metals are included to illustrate the utility of these techniques.

INTRODUCTION

THE electron microscope has achieved considerable success in the observation of objects which are sufficiently thin or which are sufficiently dispersed to allow transmission of electrons of 30–100 kv energy. Bacteria, viruses, fibers, pigments, and smokes all make suitable subjects, and a large number of papers describing electron microscopical studies of these materials have been published.

However, when efforts are made to observe "electronically opaque" specimens, as, for example, metallographic specimens, it is difficult to obtain microscopical data using the ordinary techniques. Typical metallurgical samples are far too thick to allow penetration of either light or electrons. The metallographic light microscope overcomes this difficulty by illuminating the specimen either from the side or through the objective and utilizes the light reflected by the surface to form the image. The same procedure is possible, at least in principle, with the electron microscope. Practical difficulties, however, have thus far prevented general utilization of this technique.

Another method designed to accomplish the same purpose scans a small area of the surface with a very fine electron probe and reproduces the surface on the screen of a cathode-ray tube or facsimile recorder. This method was first suggested and developed into a practical instrument by Knoll[1] and von Ardenne.[2] More recently, Zworykin, Hillier, and Snyder[3] have constructed an instrument of this type which produces images having a resolving power of approximately 500A, considerably higher than that of the best optical metallographs. However, the complexity of the equipment as well as the operating problems which were encountered have proved difficult to overcome.

The limitations of direct methods of surface observation by electron optical methods led to the development of ingenious techniques for preparing thin film replicas of the surfaces of bulk materials. H. Mahl[4] was the first to suggest this possibility in connection with electron microscopical studies. He prepared thin oxide films on aluminum surfaces and floated them off after chemically etching away the base metal. He also suggested[5] the possibility of generalizing the method by using collodion, rather than natural oxide films, to form impressions of surfaces.

A number of improved variations of this technique have appeared during the past few years. Although similar in many respects, the methods vary chiefly in being either one-step or two-step processes. The direct, one-step replica technique consists in casting a thin film of resin on the prepared surface followed by mechanical stripping under water. This method, first described by Schaefer and Harker[6] of the General Electric Laboratories, has been used by a number of workers because of the speed and simplicity of the technique.

Two-step processes have been developed to eliminate the necessity for mechanical stripping of thin films, a procedure which is somewhat difficult in any case and which is usually impossible when studying deeply etched or reentrant structures. The two-step method consists in first reproducing the prepared surface in a bulk

[1] M. Knoll, Zeits. f. tech. Physik **16**, 489 (1935).
[2] M. v. Ardenne, Zeits. f. Physik **109**, 553 (1938).
[3] V. K. Zworykin, J. Hillier, and R. L. Snyder, A.S.T.M. Bull. No. 117, 15 (1942).
[4] H. Mahl, Zeits. f. tech. Physik **21**, 17 (1940).
[5] H. Mahl, Jahrbuch AEG-Forschung **17**, 67 (1940).
[6] V. J. Schaefer and D. Harker, J. App. Phys. **13**, 427 (1942).

material which can be easily removed from the original and then making a thin replica of the first impression. The thin film is finally removed by some chemical method.

Zworykin and Ramberg[7] were the first to utilize this technique successfully in their silver-collodion method. A thick layer of metal, for example silver, is evaporated on the prepared surface by the usual techniques of vacuum evaporation and the metal film is then stripped mechanically from the original surface. A dilute solution of the final replica material, usually a collodion-amyl acetate solution, is flowed over the metal impression and allowed to dry. The metal film can then be removed by acid leaving a thin collodion replica which is examined in the usual way in a transmission electron microscope.

The most completely satisfactory replica technique to date is probably the polystyrene-silica method originated by Heidenreich and his colleagues[8] in the laboratories of the Dow Chemical Company. This is a two-step method in which the first replica is molded polystyrene, and the second thin film replica is evaporated silica. The polystyrene impression of the specimen is prepared in an ordinary molding press at a temperature of 160°C and with pressures in the range of 2000–5000 p.s.i. After cooling, the sides of the mold are sawed off and the original specimen is either lifted off or dissolved in dilute acid or alkali. The surface is then exposed to silica evaporating in vacuum from a heated spiral tungsten filament containing some quartz splinters. By dissolving the polystyrene base in some suitable solvent, this residual silica film may be utilized as a positive replica.

This replica method is somewhat more difficult in technique than the others and may be a relatively time-consuming procedure. Good contrast and resolution of the order of 100A are readily obtainable, however, and the replicas which result are easily duplicated and are relatively permanent. This method is perhaps the one most commonly employed at the present time and it has yielded outstanding results in many important metallurgical problems.

LOW PRESSURE REPLICA FORMATION

The restrictions attendant to all electron microscopical observations have been discussed at length in many places during the past few years. The action of an intense electron beam upon delicate specimens and the vacuum-induced dehydration effects are observational problems for which there as yet have been found no satisfactory solution. Consequently, the electron microscopical studies of fibers and organic crystals which have been attempted have often been seriously circumscribed.

In certain instances, partial solutions to the problem have been found. For example, in the study of cotton and wool fiber[9] it has been found that a mechanical disintegration of the material subdivides the parent fibers to such an extent that it is possible to observe the disintegrated specimen without any obvious artifacts arising from electron bombardment or vacuum desiccation.

However, in addition to obtaining information concerning the fibrillar structures, one is often desirous, for practical reasons, of observing the surface features of textile and paper fibers. Mechanical disintegration would in general destroy many of the features which might be of potential interest. This suggested the development of supplementary techniques which would allow surface examination of fibers, thereby augmenting the data obtained by the simple disintegration method.

One fairly obvious approach to the problem is to consider the possibility of reproducing the surface of the specimen in a suitable plastic. Thin replicas suitable for electron microscopical studies could then be formed by one of the several available techniques.

The chief modification necessary in an application of the polystyrene-silica replica technique to the study of organic, rather than metallic, surfaces is in a reduction of the pressure necessary to form the primary impression. Delicate organic materials would be crushed beyond all recognition, if one were to subject them to pressures of the order of 2000 p.s.i., as in the formation of metallurgical replicas. A reduction in the pressure, on the other hand, would necessarily entail a loss in resolution since the plastic would not be forced into all of the minute interstices of the surface. Nevertheless, it was decided to investigate low pressure methods in an effort to discover the possibilities and limitations of such a technique. Methods of this general nature have been successfully used in the past by

[7] V. K. Zworykin and E. G. Ramberg, J. App. Phys. **12**, 692 (1942).
[8] R. D. Heidenreich and L. A. Matheson, J. App. Phys. **15**, 423 (1944).
[9] R. B. Barnes and C. J. Burton, Ind. Eng. Chem. **35**, 120 (1943).

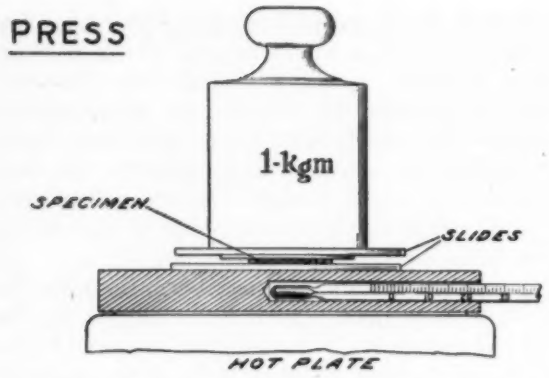

FIG. 1(a). Low pressure press used for forming initial polystyrene replica.

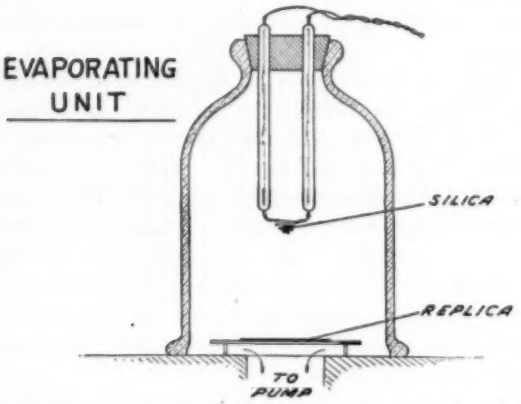

FIG. 1(b). Apparatus for vacuum evaporation of silica.

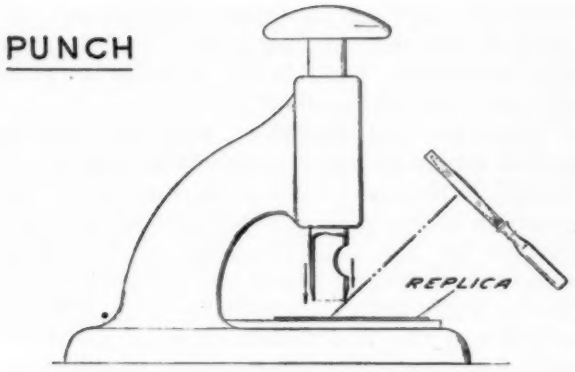

FIG. 1(c). Punch used for selecting desired portion of composite polystyrene-silica replica.

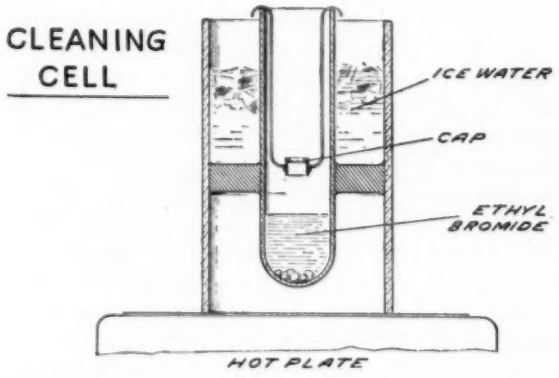

FIG. 1(d). Ethyl bromide boiler for removing polystyrene from final silica replica.

optical microscopists[10] and accordingly, the possibility of producing impressions which would have adequate resolution to allow electron microscopical study was considered. After a number of preliminary experiments, a suitable method was devised and details of the technique are illustrated in Figs. 1(a)–(d).

The object under investigation is sandwiched with a ½-inch square of 1-mil polystyrene film* between two ordinary glass microscope slides. The "sandwich" is then subjected to a pressure of 1 kilogram and is heated to approximately 160°C. The heating was originally carried out in a small oven but it has recently been found that heating on a small hot-plate yields entirely satisfactory results. After cooling, the specimen is removed from the styrene impression either by mechanical stripping or by some suitable solvent.

Obviously, polystyrene is not the only resin suitable for formation of the original impression. In fact, other thermoplastic resins with lower softening points have in some cases been found preferable. One particularly useful material is isobutyl methacrylate, a resin having a softening point below 100°C, and consequently particularly useful for forming impressions of the surfaces of water-containing substances.

A film of silica a few hundred Angstroms in thickness is deposited on the polystyrene using the ordinary technique of vacuum evaporation.[11] The desired portion of the polystyrene-silica composite may be selected and punched out with a tool which has an inside diameter slightly less than that of the ordinary specimen holder (*ca.* ⅛ inch), and the resulting disk placed in a specimen cap with a 200-mesh screen support. It has been found that the polystyrene, normally

[10] J. I. Hardy and T. M. Plitt, *Wildlife Circular 7*, U. S. Government Printing Office.

* Obtained from Plax Corporation, Hartford, Connecticut.

[11] R. D. Heidenreich and V. G. Peck, J. App. Phys. **14**, 23 (1943).

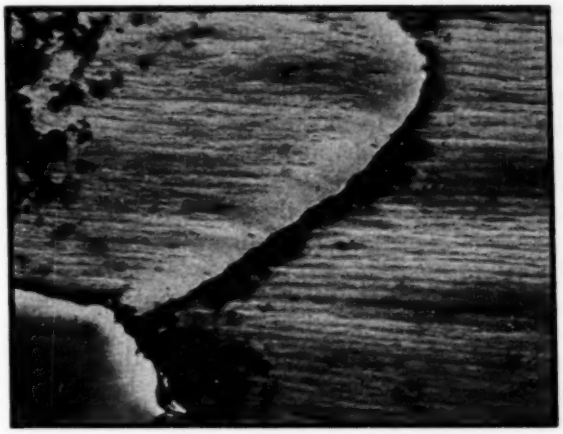

FIG. 2(a). Polystyrene-silica replica of wool fiber showing scales and longitudinal striations. 9500×.

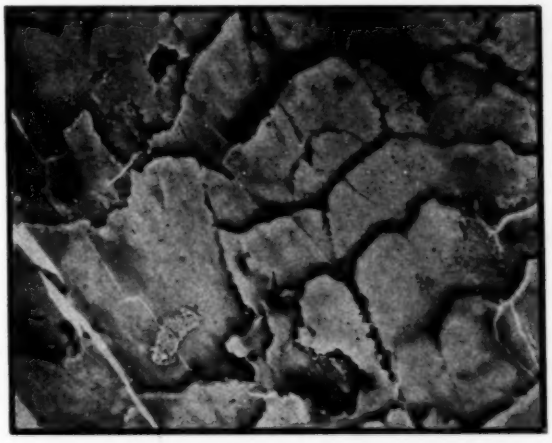

FIG. 2(b). Polystyrene-silica replica of human hair showing characteristic scale structure. 900×

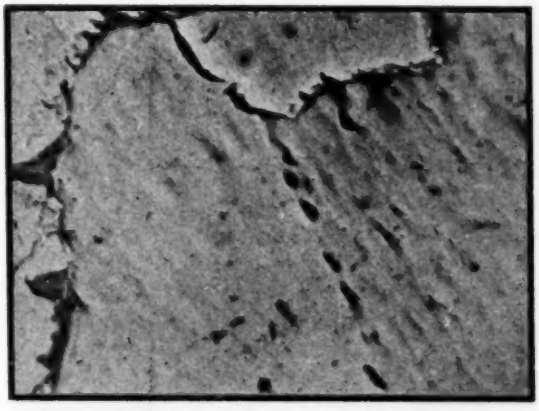

FIG. 2(c). Polystyrene-silica replica of human hair at higher magnification. Surface detail on scales is clearly visible. 9000×.

FIG. 2(d). Polystyrene-silica replica of wool fiber showing gross scale features. 1200×

dissolved by immersing the composite replica in ethyl bromide, can be somewhat more effectively removed by placing the specimen holder in the condensing vapor of the solvent.

The cell used in this work consists of an ethyl-bromide boiler and a small condenser filled with ice water. The mount is placed in a region where the vapors condense, pick up some polystyrene, and drop back into the boiling flask. This brings the replica in contact with only pure ethyl bromide and minimizes fracture of the thin silica films by liquid pressure. Furthermore, once the polystyrene is dissolved, the silica film is properly oriented in the specimen holder and is ready without further handling for immediate insertion into the electron microscope.

The first application of this new technique was made in a study of fiber surfaces. In order to show the entire width of the fibers, it was necessary to use relatively low electronic magnifications. Even at these magnifications, however, the resolving power of the electron microscope is adequate to allow considerable useful optical magnification.

Figures 2(a)–(d) show electron micrographs of a group of wool fiber replicas. The scales and longitudinal striations are characteristic of fur

FIG. 3. Polystyrene-silica replica of nylon fiber showing "die-marks." 1200×

fibers and, judging from data obtained by means of the visual microscope, the reproduction is excellent.

Surface replicas of nylon show no details of particular interest. This fiber is prepared by extruding the hot plastic through a small orifice, and it appears probable that the long grooves visible in the electron micrograph of Fig. 3 are die-marks resulting from imperfections in the spinnerette.

Cotton is a fiber which has been exhaustively studied by many research workers. One of the more important microscopical features of cotton fibers is the cross-hatch structure formed by fibrils spiralling around the fiber axis. These spirals, one right-handed and one left-handed, form an angle of approximately 57° with one another, a value which has been substantiated by x-ray diffraction studies. A glance at Figs. 4(a) and (b) clearly shows that the electron micrographs prepared from the silica replicas of cotton fibers fail to show this phenomenon, the structure shown by the replicas being more nearly reminiscent of a series of parallel single spirals. Further study, however, suggests two possible explanations. First, there may be a transparent membrane over the structure, or part of the structure, so that a plastic impression of the surface will not yield information directly correlative with the visual microscopical data. The second possibility is that a cotton fiber has only a single series of parallel filamentary spirals wound around the fiber in "barber pole" fashion. If a visual microscope is focused at a point half way through the flat, ribbon-like cotton fiber, a slightly out-of-focus image will be obtained which will give the appearance of a double spiral. The fact that the angle formed by the filamentary bundles and the edge of the fiber is approximately 30° lends considerable credence to this second theory.

Thousands of electron micrographs of bacteria have been prepared during the past few years and it is known that a number of the electron microscopes in this country are being used exclusively on problems of this type. Nevertheless, the limitations imposed by the high vacuum and electron bombardment to which the organ-

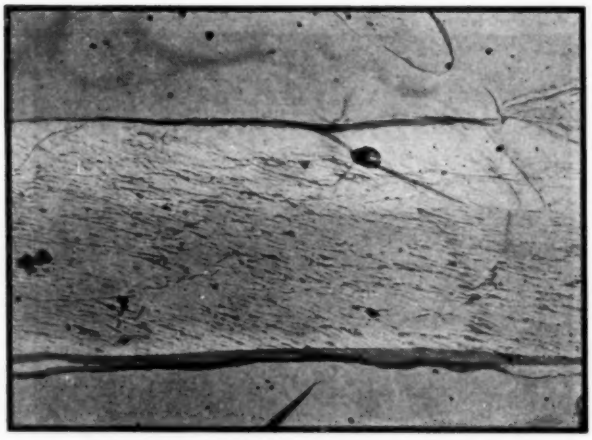

FIG. 4(a). Polystyrene-silica replica of cotton fiber showing spiral nature of surface structure. 900×.

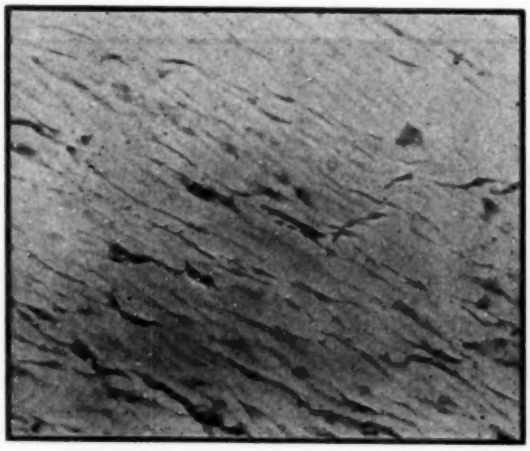

FIG. 4(b). Polystyrene-silica replica of cotton fiber showing structural features. 9000×.

isms are subjected undoubtedly introduce artifacts, although it is frequently difficult to estimate the extent to which these effects should influence interpretation of the photographs. Furthermore, the usual bacterial organisms are frequently too thick to allow adequate electron penetration so that internal morphology can be made visible.

In view of this situation, the possibility of applying the replica technique to a study of the surfaces of cells and bacterial organisms was considered. The electron micrograph of red blood cells* reproduced in Fig. 5 shows one attempt in this direction. A blood smear was covered with a styrene film, was subjected to a force of 1 kilogram, and was then heated to 150°C. After cooling, the plastic was pulled away from the smear leaving most of the cells behind. The micrograph shows the expected plate-like

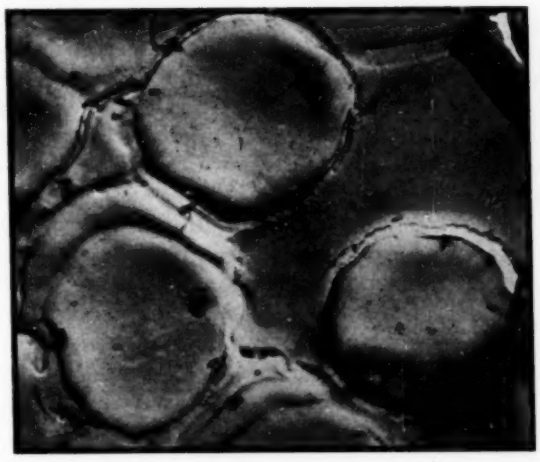

FIG. 5. Polystyrene-silica replica of red blood cells. 4000×.

appearance of the cells, together with a depression in the center. The irregular edges are probably caused by the film being forced under the cell while in the plastic state, then tearing when pulled away from the cell. Although the temperature and pressure to which the cells were subjected may seem excessively high, visual microscopical studies of preparations treated in this way fail to show any change in the cell morphology. A replica of *B. Subtilis* bacteria prepared in the same way is shown in Fig. 6.

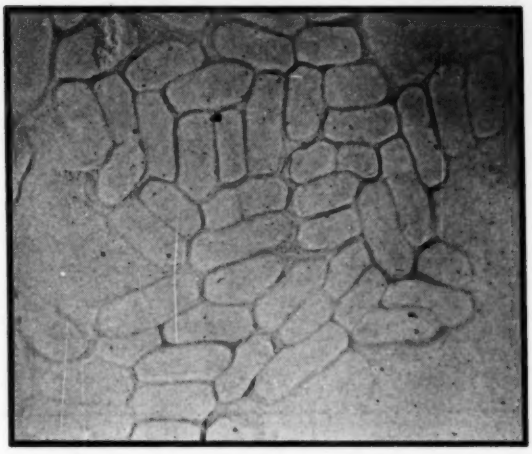

FIG. 6. Polystyrene-silica replica of *B. Subtilis* bacteria. The flagella appear as light lines throughout the field. 9000×.

Efforts were also made to study heat-sensitive organic crystals by this new replica technique. In the past, numerous attempts have been made to examine vitamins and sulfa-drugs by the ordinary methods of transmission electron microscopy. These have proved generally unsuccessful for two reasons: (1) the crystals are usually so large that nothing more than silhouette images can be obtained and (2) the intense electron bombardment causes heating effects which often destroy the crystallographic identity of the materials. The replica method under discussion minimizes these problems, and the morphology of even large crystals is clearly reproduced. Figure 7 shows a replica of sulfadiazine crystals prepared following the method already discussed at length. There are also many metallurgical

FIG. 7. Polystyrene-silica replica of sulfadiazine crystal showing crystallographic detail. 9000×.

* A successful replica of red blood cell surfaces was prepared by Mr. V. J. Schaefer of the General Electric Laboratories using the Formvar method and was described by him at the November, 1942 meeting of the Electron Microscope Society of America, Chicago, Illinois.

FIG. 8. Polystyrene-silica replica of pearlitic steel structure. 6000×

applications which may not warrant expenditure of the time and equipment required for successful utilization of the high pressure polystyrene-silica technique.[11] In such instances, the low pressure polystyrene-silica method herein described yields results which are entirely satisfactory. Figure 8 shows an electron micrograph of a surface replica of pearlitic steel prepared by the low pressure method. This micrograph clearly indicates that many metallurgical problems will be amenable to study by this method.

The use of diffraction grating replicas as a method of electron microscope magnification calibration, was first suggested several years ago in a publication[12] from these Laboratories. The simplicity and reliability of the method are well recognized by many workers in this field.[13] One rather serious drawback to more general use of this technique, however, is the fact that a relatively expensive *original* grating is required and this has in some instances proved a severe limitation. Sturdy nitrocellulose grating replicas, on the other hand, are inexpensive and widely used in spectroscopy and astronomical studies.* Accordingly, a method has been devised for using replica rather than original gratings. Silica is evaporated directly upon the ruled surface of such a grating replica. The nitrocellulose is then dissolved with acetone and the thin silica replica is ready for use in the electron microscope. Figure 9 shows a diffraction grating replica prepared in this manner. The permanence of these silica grating replicas makes them particularly suitable for frequent recalibration of the electron microscope.

[12] C. J. Burton, R. B. Barnes and T. G. Rochow, Ind. Eng. Chem. **34**, 1429 (1942).

[13] R. D. Heidenreich, J. Opt. Soc. Am. **35**, 139 (1945).

* Such replica gratings may be purchased, or may be prepared readily from an original grating, if available.

SILVER-SILICA REPLICA TECHNIQUE

Despite the success of the low pressure polystyrene-silica replica method in allowing electron microscopical observation of a number of important organic materials, certain difficulties and disadvantages of the method became apparent during the course of the work. In particular, only indifferent success resulted from efforts to obtain replicas of paper and texfile surfaces. Furthermore, many problems may be encountered which require a study of surfaces of objects which in turn have a tendency to fuse with the thermoplastic resin used for forming the impression and thus to destroy the very structures which it is desired to observe.

Accordingly, alternative methods of replica formation were considered. One method in particular seemed appealing: a combination of the Zworykin-Ramberg silver evaporation technique[7] with the silica evaporation method of Heidenreich.[14] Such a double evaporation method would at once eliminate most of the objections which had been raised to the heat and pressure requirements of the "low pressure polystyrene-silica method" described in the foregoing paragraphs. It would, of course, necessitate subjecting the original specimens to vacuum desiccation, but for many materials this would not be a serious restriction. Moreover, the mechanics of the vacuum

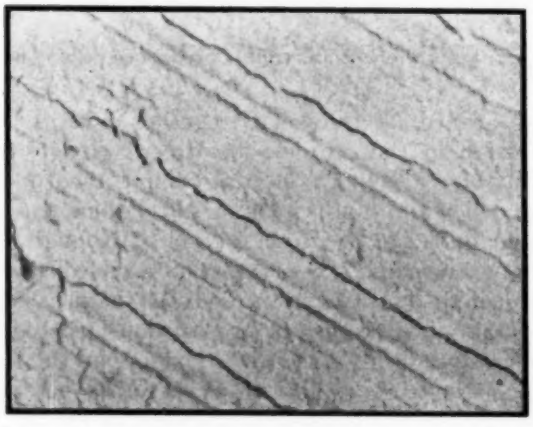

FIG. 9. Silica replica of nitrocellulose diffraction grating replica. (11,000 lines/inch.) 7500×.

[14] R. D. Heidenreich and V. G. Peck, Phys. Rev. **62**, 292 (1942).

evaporation technique should lead to higher ultimate replica resolution than with other methods where liquid surface tension effects play such an important part.

In brief, then, the method can be described in the following way. A thick coat of silver is evaporated upon the surface of the specimen. This is then removed, either by mechanical stripping or chemical solution of the specimen. The desired portion of the primary silver replica is then selected and again placed in the evaporating system where a thin coat of silica is deposited on the surface. The silver may then be removed by solution in dilute nitric acid and the silica replica examined in the usual way by means of the electron microscope.

Figure 10 shows an electron micrograph of the surface of a sheet of alpha-pulp prepared by this double evaporation method.

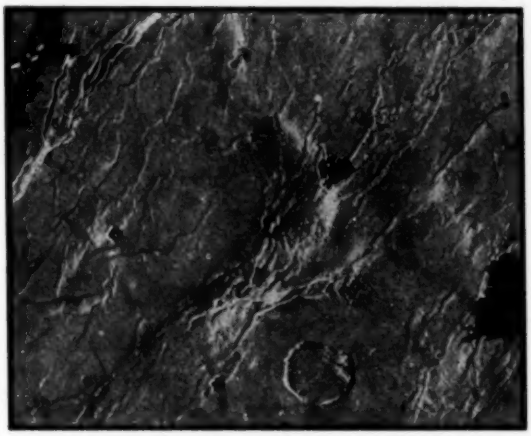

FIG. 10. Silver-silica replica of surface of alpha-pulp sheet. 900×.

Bacteriological specimens are also amenable to study by this method. A preparation of *B. pseudomonas* is shown in Fig. 11(a) and stained and unstained preparations of *Escherichia coli* are included as Figs. 11(b) and 11(c). The increase in diameter of the stained flagella is obvious.

Another problem of importance which would lend itself to this type of replica treatment is the study of paint surfaces. An electron micrograph of an enamel film is included as Fig. 12.

As an illustration of the application of the silver-silica method to metallurgical studies, Fig. 13 shows a replica of the surface of a piece of stainless steel polished by the Chrysler Super-

FIG. 11(a). Silver-silica replica of *B. pseudomonus*. 9000×.

FIG. 11(b). Silver-silica replica of unstained *Escherichia coli* preparation. 7500×.

FIG. 11(c). Silver-silica replica of *Escherichia coli* stained with Liefson's flagella stain. 7500×.

FIG. 12. Silver-silica replica of sprayed enamel film. 6000×.

finish process.* The micrograph exhibits excellent resolution and appears to be entirely satisfactory.

A polished surface of limonite ore has also been examined and the replica reproduced in Fig. 14 obtained.

INTERPRETATION OF REPLICAS

Although much has been written regarding the mechanism of replica formation and the proper methods of replica interpretation, it is obvious that a considerable amount of research is still to be done in this direction. One optical illusion frequently occurs in replica photographs which to our knowledge has not been emphasized in the literature of electron microscopy and which has often been encountered in our own work.

Thus, Fig. 15 shows two photographs of a diffraction grating replica, the only difference being that one of the positive prints has been rotated 180 degrees with respect to the other. Inspection of Fig. 15(a) clearly indicates a series of parallel grooves, whereas Fig. 15(b) certainly appears to comprise a set of parallel projections. That this is not an isolated instance can be proved by a careful examination of many of the electron micrographs appearing in this and other papers pertaining to replica studies. By rotating the photographs it is usually possible to cause the illusion of a complete reversal of topography, causing a depression to appear as a projection and *vice versa*.

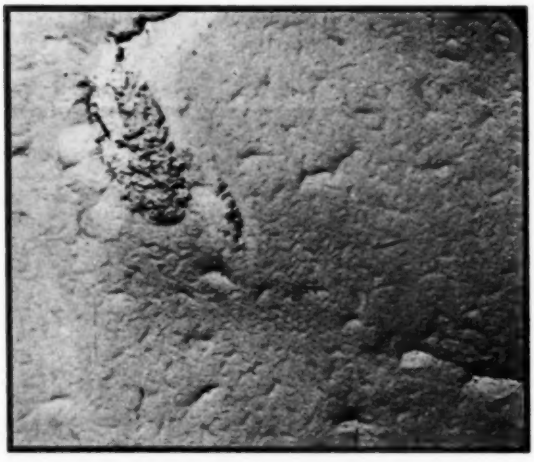

FIG. 14. Silver-silica replica of selectively polished limonite ore. 6000×.

FIG. 13. Silver-silica replica of stainless steel surface polished by Chrysler Superfinish process. 6000×.

This effect may often be seen in current pictures of the war where aerial photographs of bomb craters are at times inadvertently reversed during the printing process and the craters appear to be projecting out of the earth. Rotating the picture will ordinarily produce the normal concave appearance.

In order to illustrate the completeness of this illusion the photographs shown in Figs. 16(a) and (b) were taken. The object was a surface of

* Sample obtained from the Chrysler Corporation and photographed with their permission. The remarkable smoothness of this material is apparent from a consideration of the high magnification of the replica. The minute imperfections shown would be entirely invisible under the best visual microscope.

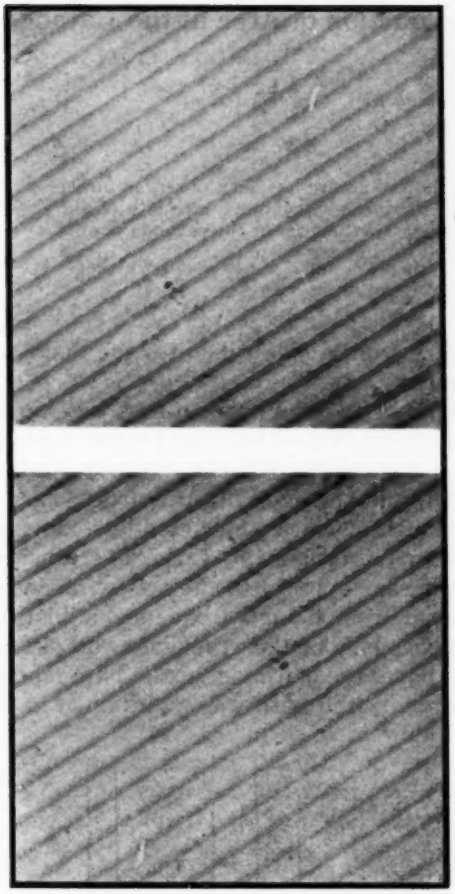

FIG. 15. Diffraction grating replica showing lines appearing as (*a*) grooves and (*b*) projections. (30,000 lines/inch.)

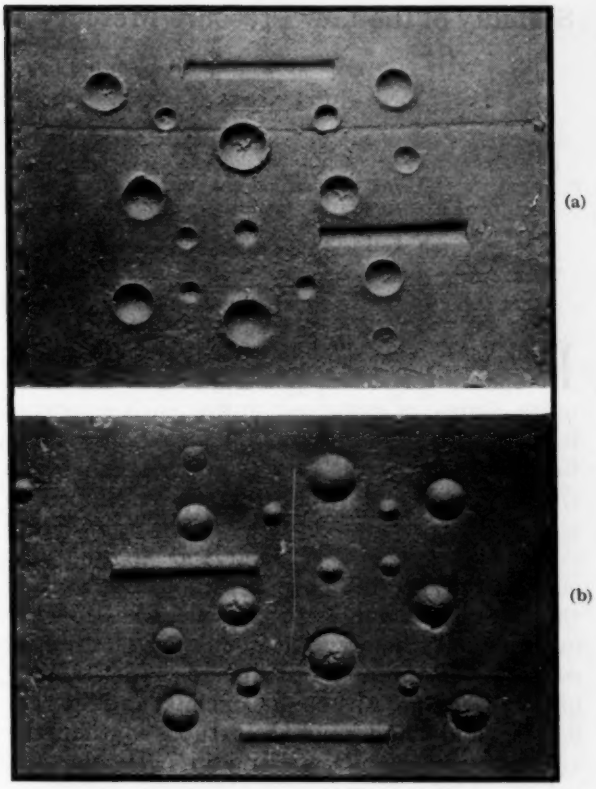

FIG. 16. Sand surface in which concave depressions have been formed. Photographic images appear concave in (*a*) and convex in (*b*), a complete topographical reversal having been effected by rotating the print through 180°.

sand in which a series of concave depressions had been made. This object was photographed vertically with a single source of illumination at approximately 45 degrees to the surface. The prints shown are identical save for orientation. A careful study of even the smallest details of these photographs clearly illustrates the phenomenon in question.

Accordingly, unless considerable care is taken in preserving the proper orientation of electron micrographs of replicas, difficulties may be encountered in interpretation. There are obviously solutions to this problem, perhaps the simplest method being the utilization of stereoscopy and contour finding.[13]

Stability of the Low Pressure Mercury Arc as a Function of Area at the Cathode Pool

CARROLL W. LUFCY AND PAUL L. COPELAND
Department of Physics, Illinois Institute of Technology, Chicago, Illinois

(Received May 21, 1945)

Using the method of Copeland and Sparing, the duration of an arc has been observed under fixed operating conditions a sufficient number of times to give a satisfactory estimate of the average life. The results show that the average life increases as the area of the cathode pool is reduced, and that the rate of change is particularly rapid for low arc currents.

PREVIOUS work[1] has indicated that an arc operating with a pool-type cathode has a finite chance of going out during each second of its operation. Conditions influencing the probability of extinction are the temperature at which the cathode pool operates,[1] the resistance in series with the arc,[2] and especially the arc current.[1] Work by Copeland and Sparing[1] has shown that conditions in the neighborhood of the cathode are primarily responsible for the extinction of the arc. The anchoring of the spot causes a very pronounced change in the stability of the discharge. Although Copeland and Sparing had several anodes in their tube so that the geometry of the plasma and in particular the length of the discharge could be varied, they failed to detect a relationship between the stability and the length of the discharge.

The method used in this investigation is basically that reported previously.[1] The arc is repeatedly started and timed until it spontaneously goes out. The average life, which is used as a quantitative measure of the stability, is obtained from a statistical analysis of the observed lives. Although the stability depends upon a great variety of operating conditions, the influence of each factor can be studied separately by varying it while maintaining all the others constant.

EXPERIMENTAL ARRANGEMENT

To obtain a convenient means for varying the cathode pool area, an experimental tube was constructed in which the mercury pool was contained within glass tubing which had been drawn down into the form of a cone as shown in Fig. 1. Since the surface area depends upon the height of the mercury in the inverted cone, it is quite easy to change the cathode pool area by distilling mercury from a side tube into the main tube or *vice versa*.

Previous work had shown that the stability of the discharge was influenced by the temperature of the bath in which the cathode pool is immersed, and the experimental arrangement contained a water bath which could be conveniently raised to surround the cathode or lowered to provide easy starting of the arc by means of the potential from a high frequency induction coil which was held against the glass wall of the tube just above the cathode pool. Throughout the experiments the bath was maintained at a tem-

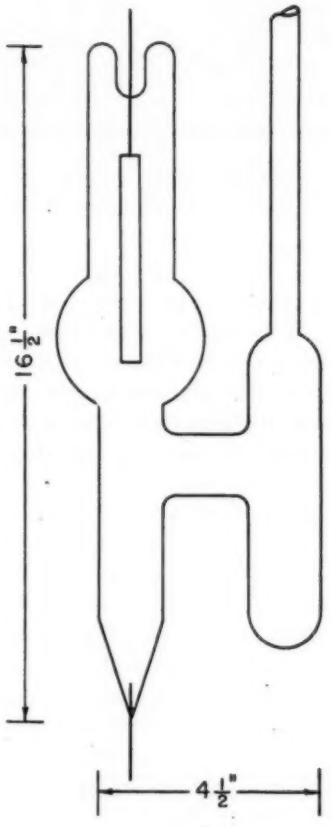

FIG. 1. Experimental arc tube.

[1] P. L. Copeland and W. H. Sparing, J. App. Phys. 16, 302 (1945).
[2] W. Kaufmann, Ann. d. Physik [4] 2, 158 (1900).

perature between 27°C and 28°C which corresponds to a vapor pressure of approximately 0.0028 mm Hg. This temperature is high enough to provide easy starting, yet when the arc is in operation it is lower than that of other parts of the tube so that the mercury is retained in the cathode pool. The temperature is sufficiently close to the one at which Copeland and Sparing obtained maximum stability for a 3.0-ampere arc that rapid variation of stability with temperature is not to be expected in this range.

Work by Ransom[3] has shown the dependence of arc stability on circuit elements. A condenser placed across the arc reduces the stability, while an inductance in series with the arc increases it. The circuits used in these experiments may be understood by referring to Fig. 2. The objectives in the design of the circuit are simplicity and reproducibility. Circuits of two types were used in the course of the work. In the first of these an effort was made to reduce the influence of unknown conditions in the d.c. mains by placing a condenser of 88 mf capacitance across the arc and a series resistance of very low inductance. In this circuit it was possible to minimize the influence of any inductance in the meter measuring the arc current simply by placing it between the condenser and the mains in series with ordinary wire wound rheostats. The voltmeter across the arc had a resistance of 20,000 ohms/volt, and it should not appreciably disturb the circuit. The resistors R_1 and R_2 directly in series with the arc across the condenser have negligible inductance. The wire coils have windings of the Ayrton-Perry type, and continuous adjustment is achieved by inserting a carbon compression rheostat R_1. With the arc in operation the condenser is maintained with a 75-volt potential difference.

To achieve greater constancy in the supply,

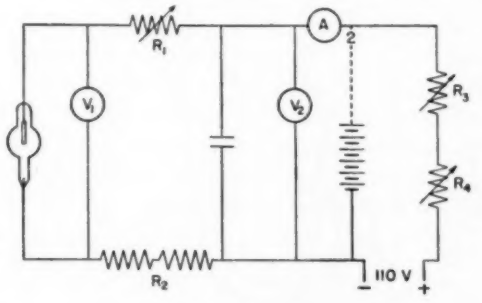

FIG. 2. Circuit diagram.

[3] R. A. Ransom, unpublished master's thesis, Department of Physics, Illinois Institute of Technology (1941).

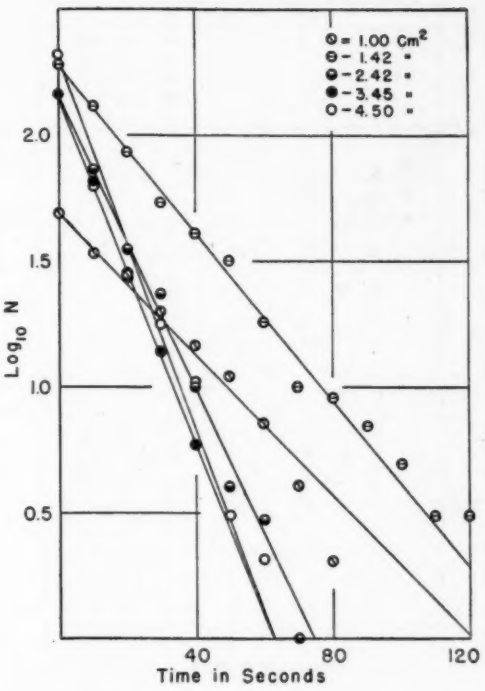

FIG. 3. Life data for 3.2-ampere arc.

a second circuit was used for most of the work. A storage battery of 89-volt electromotive force and 90 ampere-hours capacity was connected as shown at 2.

EXPERIMENTAL RESULTS

Representative results in the life tests are shown by Fig. 3, in which the logarithm of N, the number of times that the arc survived for more than t seconds, is plotted as a function of t. The arc current was 3.2 amperes. Previous work is confirmed in that the experimental points fall rather close to the straight lines drawn to represent the data. The different lines represent data for different cathode pool areas indicated by the key in the upper right-hand corner of the figure. The negative of the reciprocal of the slope for each of these lines when divided by 2.303, the ratio of natural to common logarithms, gives the average life under the conditions of operation, and there is rather obviously a regular trend of the average lives as a function of cathode pool area. Similar results were obtained at 3.6 amperes of arc current. As would be expected from previous work,[1] the average lives corresponding to the greater arc current are very much greater. A detailed examination of the data shows that in several cases the spread of the points is some-

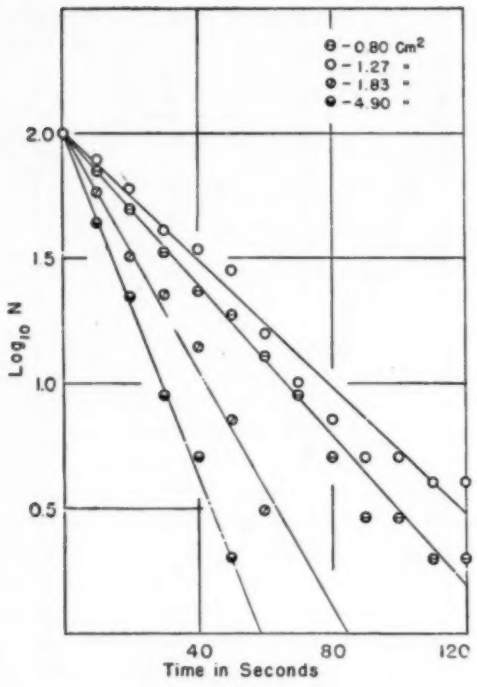

FIG. 4. Life data with battery connected.

what greater than would be expected on the basis of statistical fluctuations only. The natural interpretation would be that variation in the operating conditions was greater than it should have been. Statistical considerations[4] give the lower limit on the probable relative error to be expected in each set as $0.6745/N_0^{\frac{1}{2}}$, and since in most cases the probable error estimated from the spread of the points is quite close to that expected from fluctuation theory, the values of the probable error contained in Table I are based on statistical fluctuations only.

When the 89-volt battery was placed across the condenser, the results shown in Fig. 4 were obtained with an arc current of 3.2 amperes. The trend of the average life with cathode pool area confirms that found in Fig. 3. Similar results were obtained with the arc current maintained at 2.8, 3.0, and 3.4 amperes. Table II summarizes all these data.

The experimental tube described would not permit the use of very small cathode pool areas. For this reason a second tube, shown in Fig. 5, was constructed in which the cathode was contained in a capillary of area 0.03 cm². There was such a marked increase in the stability of the discharge for this tube that it was impractical to make average life determinations with the currents shown in Tables I and II. When this arc operates in a circuit of the first type without the storage battery connected, the average life is 8.7 sec. when the arc current is 1.7 amperes. There is an average life of 18.2 seconds for 1.8 amperes and 27.4 seconds for an arc current of 1.9 amperes.

ANALYSIS OF THE RESULTS

In view of the results presented there can be no doubt that tube geometry in the neighborhood of the cathode has a pronounced influence on the stability with which the arc operates. The natural interpretation is that when the cathode spot is free to move over the entire area of the pool the average life of the arc is a function of this area.

The results obtained by Copeland and Sparing should correlate with those obtained in the present study. A cathode pool, having an area of 23.0 cm² which is quite a bit larger than any in the present study, was used. Notwithstanding the use of higher electromotive force and in-

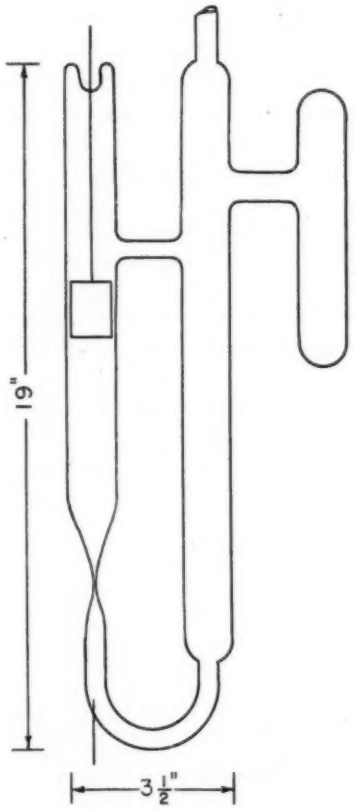

FIG. 5. Experimental tube for small cathode areas.

[4] R. Peierls, Proc. Roy. Soc. **A149**, 467 (1935).

ductive rheostats, both of which would tend to increase the stability, the observed average lives were a trifle lower than those reported here for corresponding arc currents and ambient temperatures. It would seem that anchoring the spot should be equivalent to a drastic reduction of the cathode pool area, and a spectacular increase in the average life was in fact observed when the cathode spot became anchored. These observations suggest that, when the area is large, increasing it has a relatively small effect, but that very large average lives are obtained when the cathode spot is restricted to a small pool. The relevant data are summarized in Table III. In view of the low estimated accuracy of the results, the consistency of the trends shown is very satisfactory.

A mathematical relationship in at least quali-

TABLE I. Data on stability as a function of cathode pool area. (Potential difference across condenser, 75 volts, ambient temperature (27.5±0.5)°C.)

Arc current (Amperes)	Area of cathode pool (cm²)	Average life (Seconds)	Probable relative error (Percent)
3.2	1.00	29.6	9.4
3.2	1.42	25.9	5.0
3.2	2.42	15.5	5.1
3.2	3.45	13.8	5.1
3.2	4.5	10.1	5.0
3.6	0.7	77.3	9.4
3.6	1.1	64.0	9.4
3.6	1.25	56.6	8.9
3.6	2.00	47.8	6.7
3.6	2.15	50.5	7.5
3.6	2.72	47.6	4.8
3.6	4.00	43.5	5.0
3.6	4.5	35.3	6.0

tative agreement with this behavior is

$$L = F(p, x, y, z) \exp (kI/A^{f(I)}), \qquad (1)$$

where L is the average life of the arc the current in which is I, $F(p, x, y, z)$ is some function of the vapor pressure and of other undetermined factors represented by the parameters x, y, and z. A represents the area of the cathode pool, and through $f(I)$ account is taken of the fact that the power of the pool area may depend upon the arc current. Taking the logarithm of both sides of Eq. (1) we have

$$\ln [L/F(p, x, y, z)] = (kI/A^{f(I)}). \qquad (2)$$

Repetition of the operation yields

$$\ln \ln [L/F(p, x, y, z)] = \ln k + \ln I - f(I) \ln A. \qquad (3)$$

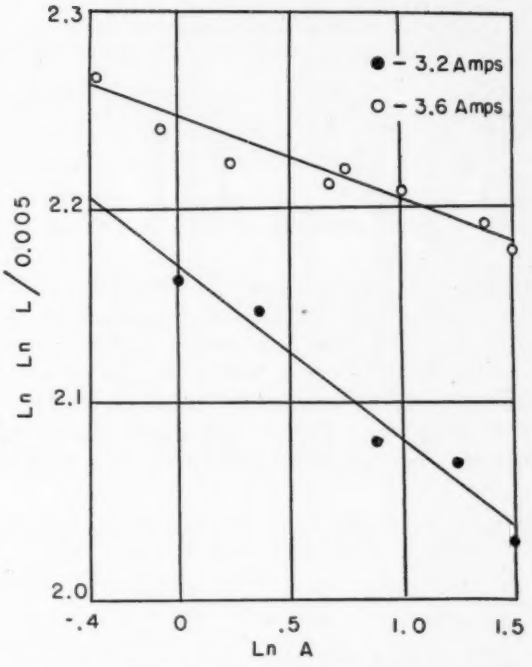

FIG. 6. Variation of arc life with cathode pool area.

In accordance with Eq. (3), it would be expected that if ln ln L were plotted against ln A in the relatively narrow range of A for which data were obtained, a linear relationship would be found. More precisely if ln ln $L/F(p, x, y, z)$ were plotted as a function of ln A, the slope of the graph should give $f(I)$ directly. Unfortunately, as is usually the case when a constant term multiplies an exponential, the value of the

TABLE II. Data on stability as a function of cathode pool area. (Electromotive force of battery, 89 volts, ambient temperature (27.5±0.5)°C, probable relative error in average life 6.7 percent.)

Arc current (Amperes)	Area of cathode pool (cm²)	Average life (Seconds)
2.8	0.65	16.5
2.8	1.25	7.2
2.8	1.95	2.8
2.8	4.9	2.25
3.0	1.2	24.6
3.0	1.86	15.4
3.0	4.9	6.2
3.2	0.8	28.2
3.2	1.27	37.3
3.2	1.83	17.7
3.2	4.9	12.3
3.4	0.9	46.3
3.4	1.25	43.0
3.4	1.85	37.0
3.4	4.9	25.4

TABLE III. Variation of constants with area of cathode constriction.

Reference	A (cm²)	K (amperes^{-1})	$F(p, x, y, z)$ (seconds)
C and S[1]	23	2.46± .06	0.011
C and L	4.9	3.78± .40	—
C and L	1.25	4.0 ±1.0	—
C and S[1]	Anchored	19.9 ±0.5	0.00096

constant $F(p, x, y, z)$ cannot be determined with accuracy that is at all satisfactory. Perhaps the best that can be done is to assign to $F(p, x, y, z)$ the value 0.005 second, which is the order of magnitude expected from the work of Copeland and Sparing. (See Fig. 6.) Although the actual values obtained for $f(I)$ depend upon $F(p, x, y, z)$, the relative placement is not changed by large errors in this constant. (Calculations have been made setting $F(p, x, y, z)=1$, and $F(p, x, y, z)=10^{-6}$. The essential conclusions to be drawn from the calculations are the same in each case.) Using $F(p, x, y, z)=0.005$, the calculations yield for $f(I)$ the values presented in Table IV.

TABLE IV. Variation of arc stability with area at the cathode pool.

Circuit type	Current (amperes)	$f(I)$
1	3.2	0.088
1	3.6	0.045
2	2.8	0.142
2	3.0	0.125
2	3.2	0.088
2	3.4	0.04

Within the limits of experimental error independent sets of data taken in different circuits support the validity of Eq. (3), and they agree in supporting the qualitative behavior of $f(I)$. The quantitative consistency happens to be better than the accuracy gives reason to expect.

CONCLUSIONS

The average life of a low pressure mercury arc with a free cathode spot is relatively large when the cathode pool area is small. Furthermore the life increases rapidly as the pool area is made very small. The variation of the average arc life with cathode pool area is rapid when the current is small and relatively slow when the current is large. This behavior may be represented within the accuracy of the present experiments by the relationship

$$L=F(p, x, y, z)\exp(kI/A^{f(I)}).$$